Forecasts and Conditionally I.I.D. Models*

Drew Fudenberg† Giacomo Lanzani‡

October 27, 2026

Abstract

We characterize when one-step-ahead forecasts are consistent with Bayesian learning about a fixed, unknown outcome distribution. With two periods of data, a conditionally i.i.d. (CIID) representation exists if and only if the matrix of joint outcome probabilities is completely positive. For binary outcomes, this is equivalent to symmetry and reinforcement (observing an outcome cannot lower the next-period probability of that outcome). For at most four outcomes, complete positivity is equivalent to positive semidefiniteness. These conditions can be tested using forecasts alone, without eliciting beliefs about the unknown distribution. Two-period forecasts cannot distinguish CIID learning from some forms of perceived persistence, but can reveal some forms of perceived reversal. For longer binary-outcome horizons, we give an explicit truncated-moment test and characterize the minimum support size of a rationalizing prior. With multiple periods and outcomes, the characterization uses count sufficiency, pairwise exchangeability, and complete positivity of the implied moment tensor.

*We thank Ian Ball, Pierpaolo Battigalli, Dan Benjamin, Simone Cerreia-Vioglio, Anujit Chakraborty, Ben Golub, Shengwu Li, Matthew Rabin, and Leeat Yariv for helpful conversations, and NSF grant SES-2417162 for financial support.

†Department of Economics, MIT

‡Department of Economics, UC Berkeley

1 Introduction

Economic models of learning typically assume that an agent learns about a fixed but unknown state, with observations independent conditional on that state. We study when one-step-ahead forecasts that are elicited conditional on alternative histories are consistent with such a *conditionally i.i.d.* (CIID) model. One-step-ahead forecasts are a natural object of study because they are often elicited in surveys, experiments, and revealed-choice settings; beliefs over full outcome sequences or latent data generating processes are much harder to observe. And as we show, such one-period forecasts can reveal that the agent is not a Bayesian who believes there is a fixed but unknown state.

Unlike much of the belief-updating literature, we do not assume that the analyst can directly elicit beliefs about the data generating process itself, only predictions of the next outcome given the outcomes so far. And unlike the exchangeability literature following de Finetti [1937], which studies when a probability measure over sequences comes from a CIID model, we do not assume the analyst can elicit beliefs about arbitrarily long or infinite sequences. Moreover, although an agent's initial beliefs may satisfy exchangeability, their updated beliefs might not.

We provide testable conditions that characterize when the forecast revisions induced by observed outcomes can be rationalized by a CIID model. With two periods and two outcomes, the characterization is simple: a CIID model exists if and only if forecasts satisfy *symmetry* and *reinforcement*. Symmetry requires that for every two outcomes i, j, the probability of j given i, multiplied by the ex-ante probability of i, is equal to the probability of i given j, multiplied by the ex-ante probability of j, which is exactly the content of exchangeability in the two-period setting. Reinforcement requires that the conditional probability of outcome i weakly increases when outcome i is observed.

When there are two periods and more than two outcomes per period, symmetry and reinforcement are still necessary but are no longer sufficient. Instead, we show that a key role is played by a square matrix M derived from the agent's forecast, whose (i, j)-th entry is the product of the ex-ante probability of the j-th outcome and the probability of the i-th outcome given the j-th outcome. When there are four or fewer outcomes, the forecast has a conditionally i.i.d. representation if and only

if M is positive semidefinite. When there are more than four outcomes, the same characterization holds, but with positive semidefiniteness replaced by the (typically stronger) condition of *complete positivity.*[1] We show that a symmetric forecast matrix is completely positive if there is a positive diagonal scaling under which the rescaled matrix is diagonally dominant.

We develop three applications of our general characterization. First, we analyze which departures from a belief in an i.i.d. process can be detected using these data. We show that it is impossible to detect beliefs that the data generating process is persistent, as in the hot-hand fallacy, because they can be rationalized by a CIID model. In contrast, some cases of the opposite bias, in which the agent believes that the first-period outcome is less likely to be realized in the next period (as in the gambler's fallacy), can be detected, as can the beliefs of a decision maker who (correctly or not) believes they have observed a garbled version of the outcome.

Second, we characterize when beliefs in a two-player incomplete-information game admit a symmetric common prior over signals, i.e., players are certain that their types are i.i.d. draws from the same distribution and share the same belief about that distribution. This assumption is standard in models of common-value auctions; it collapses to the existence of a symmetric common prior if no residual incomplete information remains given the vector of players' signals. We show that this is equivalent to a cross-player consistency condition and the complete positivity of a matrix constructed solely from players' interim beliefs about the opponent type and their ex-ante beliefs about their own type. And third, we identify the parameters of the Grether [1980] distortion of Bayesian updating model that leave it indistinguishable from a CIID representation.

When the analyst has elicited one-period-ahead forecasts over a longer time horizon T, the CIID representation imposes additional constraints: longer horizons identify higher-order moments of the latent measure and impose nontrivial consistency requirements across periods. For multi-period forecasts with binary outcomes, we derive an exact and easily checked characterization of the CIID representation by combining generalizations of the symmetry and reinforcement conditions with re-

[1]Roughly speaking, a matrix is completely positive if it can be built from a finite collection of nonnegative component vectors whose outer products add up to the matrix.

sults for the truncated Hausdorff moment problem. We also show how that Rabin, 2002's model of believers in the Law of Small Numbers can be rationalized by CIID for $T = 2$, but it necessarily fails our test for $T \geqslant 3$. Finally, we show that complete positivity of the associated moment tensor provides a general necessary and sufficient condition for multiple periods and multiple outcomes.

Related work De Finetti [1937] and Hewitt and Savage [1955] show that exchangeability is both necessary and sufficient for a probability measure on infinite sequences of random variables to correspond to learning about a fixed state: any exchangeable distribution admits a Bayesian representation as a belief over a possibly infinite collection of i.i.d. data generating processes. Thus, when the agent believes their data are exchangeable, it is as if they believe there is a fixed state of the world, whether or not they consciously think of the problem that way. Diaconis [1977] characterizes the implications of exchangeability on finite sequences, and Aldous, Ibragimov, and Jacod [2006] surveys subsequent results.

Exchangeability is an ex-ante property of the measure over sequences. A different approach asks when beliefs in two-period models before and after receiving information can be rationalized as the result of Bayesian updating. Molavi [2026] shows that beliefs about an unknown state are consistent with Bayesian updating if and only if the mean posterior belief about the state is absolutely continuous with respect to the prior. This finding generalizes the earlier work of Shmaya and Yariv [2016] by allowing an infinite state space and decision-maker subjective beliefs whose support does not match that of the true data-generating process. Bohren and Hauser [2026] characterizes the conditions under which departures from Bayesian updating can be rationalized as a consequence of Bayesian updating within a misspecified model, and Chen, Hansen, and Hansen [2020] shows how to recover a set of investor beliefs from pricing restrictions expressed in terms of conditional moments.

This paper lies at the intersection of these approaches. As in the exchangeability literature, we characterize when beliefs over observable outcomes admit a conditionally i.i.d. representation. As in the belief-revision literature, we use beliefs elicited after information is observed. The primitives, however, are neither beliefs over latent states nor beliefs over entire sequences, but finite systems of one-step-ahead forecasts.

Several other papers study related consistency conditions on beliefs. Golub and Morris [2020] characterizes the condition for having a (possibly asymmetric) common prior over signals. Sarnoff [2026] highlights that it is more common for forecasts to violate "posterior statistical sufficiency" than exchangeability.[2] Catonini and Lanzani [2026] characterizes the only form of Dutch book to which misspecified but Bayesian agents can be exposed.

The one-step-ahead forecasts we study are elicited in many field settings; see, e.g., Weber, d'Acunto, Gorodnichenko, and Coibion [2022] for surveys on beliefs about inflation and Dominitz and Manski [2011] and Greenwood and Shleifer [2014] for stock returns.[3] More broadly, this paper is also related to decision-theoretic work on the dynamic consistency of optimal plans and how they imply Bayesian updating (e.g., Epstein and Le Breton, 1993, Green and Park, 1996, and Ghirardato, 2002). Further afield, our characterization results can be used to check whether different statistical procedures lead to forecasts that are consistent Bayesian conditionally i.i.d. models, as done by Buscemi [2026] for Quantum Bayesian inference.

2 The Two-Period Model

Let $Y = \{1, ..., n\}$ be the finite set of possible outcomes. In each period $t \in \{1, 2\}$, an outcome is realized and observed. The agent's forecast of the period-1 outcome is $p \in \Delta(Y)$, and their forecast of the period-2 outcome conditional on observing outcome i is $q^{(i)} \in \Delta(Y)$. Together, we call this pair a *forecast.* We characterize when these probabilities are consistent with a conditionally i.i.d. model. In our setting, an i.i.d. model is one where outcomes are drawn independently from a fixed measure $\theta \in \Delta(Y)$. A conditionally i.i.d. model is then summarized by a probability measure $\mu \in \Delta(\Delta(Y))$.[4]

Definition 1. A forecast (p, q) has a *CIID representation* if there exists a probability

[2]Posterior statistical sufficiency means that belief at period $t + 1$ depends only on belief at period t and the period-t outcome.

[3]Also, the Michigan Survey of Consumers has a rolling panel of one-period-ahead forecasts, and only a few forecasts are recorded for each agent.

[4]For an arbitrary Borel-measurable set X in a Euclidean space, we let $\Delta(X)$ denote the probability measures on X.

measure $\mu \in \Delta(\Delta(Y))$ such that:

(i) For all $i \in Y$:

$$p_i = \int_{\Delta(Y)} \theta_i \, \mathrm{d}\mu(\theta). \tag{1}$$

(ii) For all $j \in Y$ with $p_j > 0$, the posterior measure $\mu(\cdot|j)$ is defined by Bayes' rule: for any Borel set $A \subseteq \Delta(Y)$,

$$\mu(A|j) = \frac{\int_{\theta \in A} \theta_j \, \mathrm{d}\mu(\theta)}{p_j} \tag{2}$$

and for all $i \in Y$ the conditional forecast satisfies

$$q_i^{(j)} = \int_{\Delta(Y)} \theta_i \, \mathrm{d}\mu(\theta|j). \tag{3}$$

In a CIID representation, a probability measure $\theta \in \Delta(Y)$ represents the unknown but stable data generating process: outcomes are drawn independently from θ, and μ is the agent's uncertainty about what θ is. Thus the initial probability p_i is the expected value of the latent parameter θ_i, and the period-2 probability of outcome i conditional on the first outcome being j is the expected value of θ_i conditional on seeing outcome j.

Conditionally i.i.d. models generate exchangeable sequences of observations; this implies that forecasts must be symmetric in the following sense.

Definition 2. Forecast (p, q) satisfies *symmetry* if for all $i, j \in Y$, $p_i q_j^{(i)} = p_j q_i^{(j)}$.

This follows from exchangeability: the probability of sequence (i, j) must equal that of (j, i). Note that symmetry also implies the martingale property, i.e., that $p_i = \sum_{j \in Y} p_j q_i^{(j)}$.

The following condition also holds in any conditionally i.i.d. model.

Definition 3. (p, q) satisfies *reinforcement* if $p_i \leqslant q_i^{(i)}$ for all $i \in Y$.

Reinforcement requires that the probability of observing an outcome in the next period is not decreased by observing that outcome in the current period.[5] This distin-

[5]Note that CIID representations do not imply that seeing a particular outcome makes all other outcomes less likely. For example, suppose there are three outcomes and that $\mu = (.1\theta', .9\theta'')$, with

guishes CIID models from other exchangeable models, such as the sampling without replacement examples discussed in Diaconis [1977], where reinforcement fails because each draw reduces the proportion of that outcome in the remaining population. The case where the agent is certain that they know the distribution of y and thus treats the samples as i.i.d. is the edge case of an "infinite urn" where the reinforcement inequality becomes an equality for all outcomes i.

CIID models imply reinforcement because any θ that assigns relatively high probability to today's observed outcome will assign relatively high probability to that outcome tomorrow as well.

Lemma 1. *If (p, q) has a CIID representation, then (p, q) satisfies symmetry and reinforcement.*

Proof. As noted above, symmetry follows from the fact that conditionally i.i.d. models are exchangeable. Reinforcement trivially holds if $p_i = 0$. If $p_i > 0$, we have

$$q_i^{(i)} = \frac{\int_{\Delta(Y)} \theta_i^2 \mathrm{d}\mu(\theta)}{p_i} \geqslant \frac{\left(\int_{\Delta(Y)} \theta_i \mathrm{d}\mu(\theta)\right)^2}{p_i} = p_i,$$

where the first equality follows from the fact that $p_i q_i^{(i)}$ is the probability that outcome i occurs twice in a row, the inequality follows from Jensen's inequality, and the last equality follows from the definition of p_i. □

Section 7 below shows that symmetry and reinforcement remain necessary conditions for CIID models when forecasts are elicited over more than two periods.

3 Binary outcomes

We begin by providing a simple characterization in the case of binary outcomes (i.e., Bernoulli random variables), as in de Finetti [1937]. In this case, symmetry and reinforcement are sufficient as well as necessary for the existence of a CIID representation.

$\theta' = (.4, .4, .2)$ and $\theta'' = (.1, .1, .8)$. Then $q_2^{(1)} = 5/26 > .13 = p_2$.

Theorem 1. *For binary outcomes ($n = 2$), (p, q) has a CIID representation if and only if (p, q) satisfies symmetry and reinforcement.*

Necessity follows from Lemma 1. We prove sufficiency constructively. With two outcomes, reinforcement implies $p_1 \geqslant q_1^{(2)}$. When $p_1 = q_1^{(2)}$, the unique rationalizing prior is δ_p. When $q_1^{(2)} = 0$ and $0 < p_1 < 1$, the unique prior is $p_1\delta_{(1,0)} + (1 - p_1)\delta_{(0,1)}$: the agent is uncertain between two deterministic distributions, and the first observation resolves that uncertainty. When $p_1 \in \{0, 1\}$, the unique prior is δ_p, with forecasts after the null outcome unrestricted. When $1 > p_1 > q_1^{(2)} > 0$, we show there is a CIID representation with Beta parameters

$$\alpha = \frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}} \qquad \text{and} \qquad \beta = \frac{q_1^{(2)}(1 - p_1)}{p_1 - q_1^{(2)}}. \tag{4}$$

Remark. The Beta prior is not the unique CIID representation. Indeed, Section 4.2 shows that whenever $0 < q_1^{(2)} < p_1 < 1$, there are infinitely many CIID representations, including infinitely many with finite support. The Beta prior is convenient here because it is the conjugate prior for Bernoulli outcomes, has two parameters, and yields simple moment formulas. But the key point here is that *some* CIID representation exists, not that it must be Beta.

4 General Characterization: Complete Positivity

4.1 Necessary conditions

Paralleling the development of the characterization of exchangeability provided by Hewitt and Savage [1955], we now move beyond the case of binary outcomes. We begin with a necessary condition that lets us demonstrate that reinforcement and symmetry are no longer sufficient when there are more than two outcomes. Define the $n \times n$ *forecast matrix* $M(p, q)$ by

$$m_{ij}(p, q) = p_j \, q_i^{(j)} \qquad \forall i, j \in Y.$$

By construction, $m_{ij}(p,q)$ is the joint probability of observing j in period 1 and i in period 2. For any (p,q), the associated matrix $M(p,q)$ automatically satisfies $\sum_{i\in Y}\sum_{j\in Y} m_{ij}(p,q) = 1$, since $p \in \Delta(Y)$ and each $q^{(j)} \in \Delta(Y)$. This will be important below.

The next lemma says that in a CIID model, this matrix must be positive semidefinite.

Lemma 2. *$\mu \in \Delta(\Delta(Y))$ is a CIID representation for (p,q) if and only if*

$$m_{ij}(p,q) = \int_{\Delta(Y)} \theta_i \theta_j \, d\mu(\theta) \qquad \forall i,j \in Y. \tag{5}$$

In this case, $M(p,q)$ is positive semidefinite.

The proof is based on the observation that if (p,q) has a CIID measure μ, $M(p,q)$ is the matrix of that measure's second moments.

Remark. When $|Y| = 2$, symmetry and reinforcement imply that the forecast matrix $M(p,q)$ is positive semidefinite: symmetry implies that $M(p,q)$ is symmetric, while reinforcement implies $q_2^{(1)} \leqslant p_2$ and $q_1^{(2)} \leqslant p_1$, so that

$$\det M(p,q) = p_1 p_2 \big(q_1^{(1)} q_2^{(2)} - q_1^{(2)} q_2^{(1)}\big) = p_1 p_2 \big(1 - q_1^{(2)} - q_2^{(1)}\big) \geqslant 0.$$

Since the diagonal entries of $M(p,q)$ are nonnegative, it follows that $M(p,q)$ is positive semidefinite (see, e.g., Theorem 2.4 in Berman and Shaked-Monderer [2003]).

Now consider the following example.

Example 1. *Suppose $n = 3$, the initial forecast is $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and the second-period forecasts are $q^{(1)} = (0.4, 0.5, 0.1), q^{(2)} = (0.5, 0.4, 0.1)$, and $q^{(3)} = (0.1, 0.1, 0.8)$. Clearly (p,q) satisfies symmetry and reinforcement with*

$$M(p,q) = \begin{bmatrix} 4/30 & 5/30 & 1/30 \\ 5/30 & 4/30 & 1/30 \\ 1/30 & 1/30 & 8/30 \end{bmatrix}.$$

Let $e = (1,-1,0)^\top$. Because $M(p,q)\cdot e = -\frac{1}{30}e$, $M(p,q)$ is not positive semidefinite. Therefore (p,q) does not have a CIID representation, even though it satisfies

symmetry and reinforcement.

Imagine a voter assessing a politician who can produce one of three policy outcomes: Left, Center, and Right. The voter's forecasts in the example imply that observing Left makes Center the most likely policy next period, while observing Center makes Left the most likely outcome. Although the example satisfies symmetry and reinforcement, the fact that Center and Left each boost the other the most makes the matrix $M(p,q)$ not positive semidefinite, so the forecasts cannot be reconciled with a model of learning about a politician with a stable ideological "type." Section 4.3 elaborates on when similar cycles are compatible or incompatible with a CIID representation.

4.2 General Characterization: Complete Positivity

The complete positivity of $M(p,q)$ will play a central role in our characterization.

Definition 4. The matrix M is *completely positive* if $M = BB^\top$ for some $n \times r$ nonnegative matrix B. When M is completely positive, its *cp rank* $\text{cpr}(M)$ is the smallest r for which such a B exists.

The decomposition of a completely positive matrix resembles the diagonalization of a symmetric matrix, with the key difference that the b vectors are not orthogonal. Complete positivity implies that $M(p,q)$ is positive semidefinite and therefore symmetric. Since symmetry of $M(p,q)$ is equivalent to symmetry of the forecast, complete positivity implies symmetry, and we will soon see that it also implies reinforcement. There is an extensive literature on completely positive matrices, see, e.g., Berman and Shaked-Monderer [2003].

The next result shows that $M(p,q)$ encodes all the restrictions implied by CIID representations in two-period models.

Theorem 2. *For a forecast (p,q), the following are equivalent:*

1. *$M(p,q)$ is completely positive.*
2. *(p,q) has a CIID representation.*
3. *(p,q) has a CIID representation that has finite support.*

To prove the result, we introduce a strengthening of complete positivity called simplex complete positivity, and show that it is equivalent to complete positivity in our setting because the entries of any $M(p,q)$ sum up to one. An $n \times n$ matrix M is *simplex completely positive* if it admits the decomposition

$$M = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)} \pi^{(s)\top}, \quad \gamma_s > 0, \; \sum_{s=1}^{r} \gamma_s = 1, \; \pi^{(s)} \in \Delta(Y) \tag{6}$$

for some integer r. It is immediate that a simplex completely positive matrix is completely positive. When $\sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} = 1$, the converse is also true.

Claim 1. *If M is completely positive and $\sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} = 1$, then M is simplex completely positive.*

The proof of Theorem 2 uses this claim to establish the cycle of implications (2) $\Rightarrow$ (1) $\Rightarrow$ (3) $\Rightarrow$ (2). To show that (2) implies (1), we note that if μ is a CIID representation for (p,q), then by Lemma 2, $M(p,q)$ is a mixture of rank-one matrices $\theta\theta^\top$ with $\theta \in \Delta(Y)$. From Carathéodory's theorem, it can be written as a finite convex combination of them, so it is completely positive. The proof that (1) implies (3) uses Claim 1 to infer that $M(p,q)$ is simplex completely positive so that $M(p,q) = \sum_{s=1}^{r} \gamma_s \pi^{(s)} \pi^{(s)\top}$ for some $\pi^{(1)}, \ldots, \pi^{(r)} \in \Delta(Y)$ and weights $\gamma_s > 0$ with $\sum_{s=1}^{r} \gamma_s = 1$. Define the discrete measure $\mu := \sum_{s=1}^{r} \gamma_s \delta_{\pi^{(s)}}$, where δ is the Dirac measure. By construction, this measure has second moments $\int_{\Delta(Y)} \theta_i \theta_j \, \mathrm{d}\mu = m_{ij}(p,q)$, so by Lemma 2 it is a CIID representation with finite support. Finally, that (3) implies (2) is trivial. Note that combining this theorem with Lemma 1 shows that if $M(p,q)$ is completely positive, then (p,q) satisfies reinforcement.

Non-Uniqueness Theorem 2 has implications for the uniqueness of CIID representations. Say that a forecast (p,q) is *dogmatic* if $p = q^{(i)}$ for all $i \in Y$. Dogmatic forecasts clearly admit a unique CIID representation, namely the Dirac measure on p. The next corollary gives two simple conditions under which this uniqueness necessarily fails.

Corollary 1. *Let p be strictly positive. If there is a CIID representation for (p,q) with infinite support, then the Bayes-rationalizing model is not unique. In particular,*

with binary outcomes, any nondogmatic forecast with $q_1^{(2)} > 0$ *that admits a CIID representation has at least two of them.*

A CIID representation allows the agent to be uncertain over an arbitrarily rich, even uncountable, collection of possible data generating processes. The next result shows that for any desired accuracy, the forecast matrix generated by such a prior can be approximated by one generated by a prior with finite support. The required support size is independent of the number of outcomes, and grows only as the inverse square of the approximation error. The theorem also implies that a forecast matrix near the completely positive cone must be close to a forecast (p', q') admitting a CIID representation with small finite support: the initial probabilities are close, and the conditional forecasts are close when their errors are weighted by the original outcome probabilities. If those probabilities are bounded away from zero, every conditional forecast is close. These approximation bounds on forecasts depend on the number of outcomes and, for the uniform conditional bound, on the lower bound on initial probabilities. This is especially relevant when forecast probabilities are estimated with sampling error or reported with noise. We write $\|A\|_F = (\sum_{i=1}^n \sum_{j=1}^n A_{ij}^2)^{1/2}$ for the Frobenius norm of an $n \times n$ matrix.

Theorem 3. *Suppose matrix* $M(p, q)$ *has Frobenius distance at most* $\varepsilon > 0$ *from the completely positive cone. There exists a forecast* (p', q') *admitting a CIID representation with at most* $\lceil \varepsilon^{-2} \rceil$ *atoms such that*

$$\|M(p', q') - M(p, q)\|_F \leqslant (n + 2)\varepsilon, \tag{7}$$

and

$$\begin{aligned} \|p' - p\|_1 &\leqslant n(n + 2)\varepsilon, \\ \sum_{j=1}^n p_j \|q'^{(j)} - q^{(j)}\|_1 &\leqslant 2n(n + 2)\varepsilon. \end{aligned} \tag{8}$$

Moreover, if p *is strictly positive, the same forecast satisfies*

$$\max_{1 \leqslant j \leqslant n} \|q'^{(j)} - q^{(j)}\|_1 \leqslant \frac{2\sqrt{n}\,(n + 2)}{\min_{j \in \{1,\ldots,n\}} p_j}\,\varepsilon. \tag{9}$$

The proof finds a completely positive matrix $\bar{A}$ close to $M(p, q)$, normalizes it

so its entries sum to one (using the fact that $M(p,q)$ already sums to one, so normalization changes $\bar{A}$ by at most $n\varepsilon$), and then applies Lemma A.1 to approximate $\bar{A}$ by the second-moment matrix of a $\lceil \varepsilon^{-2} \rceil$-atom CIID prior. The forecast (p', q') is recovered from this matrix by reading off column sums and ratios, and the bounds on forecast errors follow from Cauchy–Schwarz applied column by column.

The approximation result has a converse with constants independent of the number of outcomes. To state it, let $\mathcal{CP}_n$ denote the cone of completely positive $n \times n$ matrices.

Theorem 4. *Let (p,q) be a forecast and let (p',q') be a forecast admitting a CIID representation. Then*

$$\inf_{A \in \mathcal{CP}_n} \|M(p,q) - A\|_F \leqslant \|p - p'\|_2 + \frac{1}{\sqrt{2}} \left(\sum_{j=1}^{n} p_j^2 \|q^{(j)} - q'^{(j)}\|_1^2 \right)^{1/2} \leqslant \frac{\|p - p'\|_1 + \sum_{j=1}^{n} p_j \|q^{(j)} - q'^{(j)}\|_1}{\sqrt{2}}. \tag{10}$$

In particular, if $p' = p$, then

$$\inf_{A \in \mathcal{CP}_n} \|M(p,q) - A\|_F \leqslant \frac{1}{\sqrt{2}} \left(\sum_{j=1}^{n} p_j^2 \|q^{(j)} - q'^{(j)}\|_1^2 \right)^{1/2}. \tag{11}$$

The next corollary uses this theorem to show that if the agent's posterior beliefs are close in total variation distance to the Bayesian posteriors, then the forecast matrix must lie close to the completely positive cone. Moreover, the bound does not depend on the size of the outcome space. Its proof is in Online Appendix B.2.

Corollary 2. *Suppose there is a prior $\mu \in \Delta(\Delta(Y))$ such that $p = \int_{\Delta(Y)} \theta \, \mathrm{d}\mu(\theta)$. Suppose also the agent's belief after j is $\widetilde{\mu}_j$ and $q^{(j)} = \int_{\Delta(Y)} \theta \, \mathrm{d}\widetilde{\mu}_j(\theta)$. Let $\eta_j = d_{\mathrm{TV}}(\widetilde{\mu}_j, \mu(\cdot|j))$ for $p_j > 0$, and set $\eta_j = 0$ for $p_j = 0$. Then*

$$\inf_{A \in \mathcal{CP}_n} \|M(p,q) - A\|_F \leqslant \sqrt{2} \left(\sum_{j=1}^{n} p_j^2 \eta_j^2 \right)^{1/2} \leqslant \sqrt{2} \sum_{j=1}^{n} p_j \eta_j. \tag{12}$$

Small Number of Outcomes Maxfield and Minc [1962] showed that for $n \leqslant 4$ a matrix is completely positive if and only if it is nonnegative and positive semidefinite. Thus an immediate corollary of Theorem 2 is that for four or fewer outcomes, positive

semidefiniteness of the matrix $M(p,q)$ captures all of the empirical implications of conditionally i.i.d. models. This result provides a computationally simple test.

Corollary 3. *When $n \leqslant 4$, the following are equivalent:*

1. *(p,q) has a CIID representation.*

2. *$M(p,q)$ is positive semidefinite.*

The next example shows that the equivalence of (1) and (2) does not extend to $n > 4$ even when (p,q) satisfies reinforcement.

Example 2. *Let $n = 5$, $p = \left(\frac{3}{23}, \frac{4}{23}, \frac{4}{23}, \frac{4}{23}, \frac{8}{23}\right)$, $q^{(1)} = \left(\frac{1}{3}, \frac{1}{3}, 0, 0, \frac{1}{3}\right)$, $q^{(2)} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0, 0\right)$, $q^{(3)} = \left(0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0\right)$, $q^{(4)} = \left(0, 0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$, and $q^{(5)} = \left(\frac{1}{8}, 0, 0, \frac{1}{8}, \frac{3}{4}\right)$. It is immediate that (p,q) satisfies reinforcement and symmetry and that*

$$M(p,q) = \begin{bmatrix} 1/23 & 1/23 & 0 & 0 & 1/23 \\ 1/23 & 2/23 & 1/23 & 0 & 0 \\ 0 & 1/23 & 2/23 & 1/23 & 0 \\ 0 & 0 & 1/23 & 2/23 & 1/23 \\ 1/23 & 0 & 0 & 1/23 & 6/23 \end{bmatrix}.$$

Here $M(p,q) = A/23$, where A is the matrix from Example 2.4 of Berman and Shaked-Monderer [2003]. Because positive semidefiniteness and complete positivity are both preserved under positive scaling (see Theorem 2.2 in Berman and Shaked-Monderer, 2003), $M(p,q)$ is positive semidefinite but not completely positive. By Theorem 2, (p,q) does not admit a CIID representation.

4.3 Cycles

We can gain further insight into the CIID model by analyzing its implications for cyclical patterns in belief updating. Throughout this subsection, assume that $p_i > 0$ for every $i \in Y$. The direction of belief updates is determined by the covariance structure of the distribution of θ, because if forecast (p,q) admits a CIID representation μ, then

$$q_j^{(i)} = \frac{\mathbb{E}_\mu[\theta_j \theta_i]}{\mathbb{E}_\mu[\theta_i]} = p_j + \frac{\mathrm{Cov}_\mu(\theta_j, \theta_i)}{p_i} \qquad \forall i, j \in Y. \tag{13}$$

Thus, observing outcome i makes outcome j strictly more likely if and only if $\mathrm{Cov}_\mu(\theta_j, \theta_i) > 0$.[6] The simplex constraint $\sum_{k \in Y} \theta_k = 1$ implies

$$0 = \mathrm{Var}_\mu\Big(\sum_{k \in Y} \theta_k\Big) = \sum_{i \in Y} \mathrm{Var}_\mu(\theta_i) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}_\mu(\theta_i, \theta_j). \tag{14}$$

Since variances are nonnegative, this identity implies that the sum of all pairwise covariances must be nonpositive. Intuitively, there is a fixed "covariance budget": any positive covariances must be offset by sufficiently negative covariances elsewhere.

Definition 5. A forecast (p, q) exhibits a *full cycle* if there is an ordering of outcomes such that $q^{(i)}_{i+1} \geqslant p_{i+1}$ for all $i \in Y$, where indices are taken modulo n, and the inequality is strict for at least one i.

A full cycle requires n nonnegative covariances arranged along a cyclic permutation of all outcomes with at least one covariance strictly positive. This requires that the positive covariances are offset by sufficiently negative covariances elsewhere.

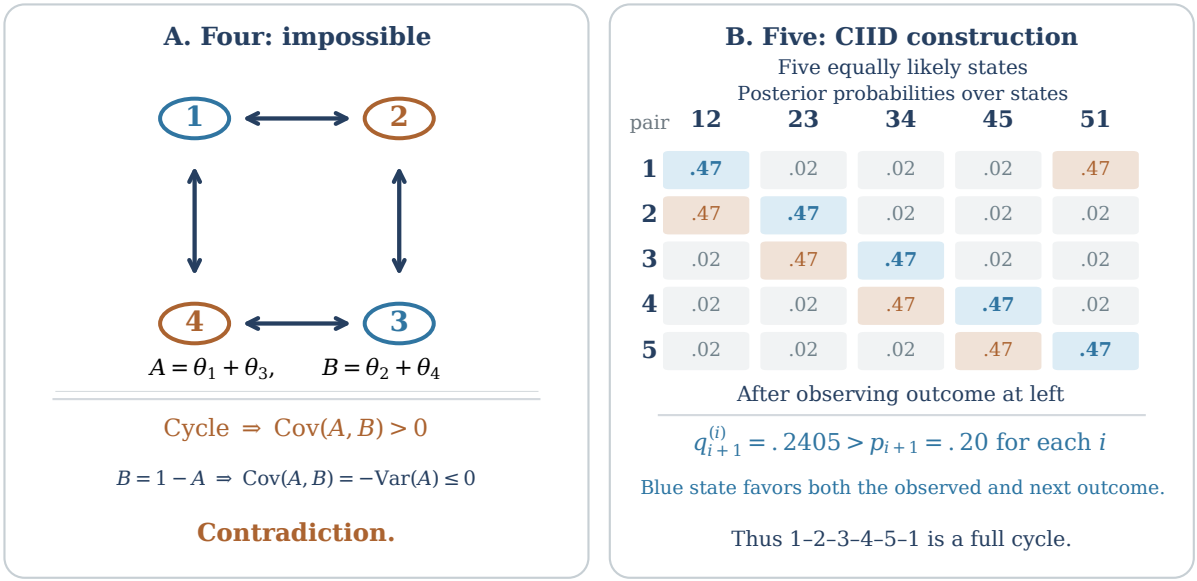
Proposition 1. *Let $n \leqslant 4$. If forecast (p, q) admits a CIID representation, then it does not exhibit a full cycle.*

Panel A of Figure 1 illustrates why a full cycle is impossible with four outcomes. Partition the outcomes into the opposite pairs $\{1, 3\}$ and $\{2, 4\}$. A full cycle would make all four covariances across these pairs nonnegative, with at least one strictly positive, so $\theta_1 + \theta_3$ and $\theta_2 + \theta_4$ would be positively correlated. But these two sums add to one, so their covariance is $-\mathrm{Var}_\mu(\theta_1 + \theta_3) \leqslant 0$.

Remark (Short cycles)**.** Proposition 1 rules out full cycles when $n \leqslant 4$, i.e., configurations in which there exists an ordering of all outcomes such that $q^{(i)}_{i+1} \geqslant p_{i+1}$ for every i (indices modulo n) with some strict inequality. It does not rule out shorter cycles involving only a proper subset of outcomes: the simplex constraint $\sum_{i \in Y} \theta_i = 1$ forces only the sum of all pairwise covariances to be nonpositive, so the positive covariances required by a shorter cycle can be offset by sufficiently negative covariances involving the remaining outcomes.

The simplex constraint is less binding for larger n, as the necessary negative

[6] So observing an outcome makes all the others less likely if and only if every covariance is negative.



Panel A. Each arrow means that observing one outcome does not lower the probability of the other; at least one increase is strict. The four edges would therefore make $\theta_1+\theta_3$ and $\theta_2+\theta_4$ positively correlated. But these sums add to one.

Panel B. Five equally likely states each assign probability .47 to an adjacent pair and .02 to other outcomes. Rows show posterior probabilities over states. After observing outcome i, its successor's probability rises from .20 to .2405.

Figure 1: Full cycles with four and five outcomes.

covariances can be assigned to non-adjacent pairs, leaving the cycle covariances free to be positive.[7]

Proposition 2. *If $n \geqslant 5$, there exists a forecast (p, q) that exhibits a full cycle and has a CIID representation.*

Panel B of Figure 1 illustrates the construction for five outcomes. Observing any outcome raises the probability of its successor, generating a full cycle through Bayesian learning about a fixed state.

4.4 Scaled Diagonal Dominance

Complete positivity is both necessary and sufficient for a Bayes-rationalizing conditionally i.i.d. representation, but may be difficult to verify directly. This motivates

[7] It is not a coincidence that the same critical value of n appears here as in the relation between complete positivity and PSD: In both cases the issue is that the geometry of $\mathbb{R}^n$ for $n \geqslant 5$ is qualitatively different (cf. Berman and Shaked-Monderer, 2003).

the search for easier-to-verify sufficient conditions for complete positivity.

Definition 6. Forecast (p, q) satisfies *scaled diagonal dominance* if there exist weights $s \in \mathbb{R}^Y_{++}$ with $\sum_{i \in Y} s_i = 1$ such that $p_i s_i q_i^{(i)} \geqslant \sum_{j \in Y \setminus \{i\}} p_j s_j q_i^{(j)}$ for all $i \in Y$.

The next corollary shows that scaled diagonal dominance is sufficient for a CIID representation, and provides an easily computed criterion under which this condition holds. Let $M(p, q)$ satisfy $m_{ii}(p, q) > 0$ for all $i \in Y$, and define the *row–ratio* matrix $R(p, q)$ by

$$r_{ij}(p, q) = \begin{cases} \dfrac{m_{ij}(p, q)}{m_{ii}(p, q)}, & i \neq j, \\ 0, & i = j. \end{cases}$$

Corollary 4 (Sufficient conditions for CIID rationalizability)**.** *Let (p, q) be a forecast that satisfies symmetry.*

(i) If (p, q) satisfies scaled diagonal dominance, then it admits a CIID representation.

(ii) If $p_i q_i^{(i)} > 0$ for all $i \in Y$ and the spectral radius of $R(p, q)$ is less than 1, then (p, q) satisfies scaled diagonal dominance and hence admits a CIID representation.

Remark. Let $\|A\|_\infty := \max_{i \in Y} \sum_{j \in Y} |a_{ij}|$ denote the induced matrix infinity norm. Strict row diagonal dominance of $M(p, q)$ is equivalent to $\|R\|_\infty < 1$. Since $\rho(R) \leqslant \|R\|_\infty$, the spectral condition in part (ii) is strictly weaker; see Example 5 in Online Appendix B.1.

5 Non-Uniqueness of CIID Representations

When a forecast has a CIID representation, it need not be unique. Moreover, uniqueness can fail in two ways: there may be CIID representations with different support sizes, and there may be multiple CIID representations with the same support size. This follows from the fact that CIID forecasts are determined by the first two moments of the distribution of θ, and many distributions can match those moments while differing at higher moments. This section illustrates this point for binary outcomes, and then discusses what holds more generally.

5.1 Non-uniqueness with Binary Outcomes

Suppose there are two outcomes $Y = \{1, 2\}$, and suppose forecast (p, q) has a non-dogmatic CIID representation μ. Now let $0 \leqslant \theta_L < \theta_H \leqslant 1$, fix $\lambda \in (0, 1)$ and let μ_λ be the two-point prior placing mass λ on θ_L and $1 - \lambda$ on θ_H.[8] Let $h := \theta_H - \theta_L > 0$, then $\mathbb{E}_{\mu_\lambda}[\theta_1] = \mathbb{E}_\mu[\theta_1]$ is satisfied if and only if $\theta_L = p_1 - (1-\lambda)h$ and $\theta_H = p_1 + \lambda h$. Moreover, $\mathbb{E}_{\mu_\lambda}[\theta_1^2] = \mathbb{E}_\mu[\theta_1^2]$ is satisfied if and only if $h^2 = \dfrac{\mathbb{E}_\mu[\theta_1^2] - (\mathbb{E}_\mu[\theta_1])^2}{\lambda(1-\lambda)}$, so the two-point priors that match $(\mathbb{E}_\mu[\theta_1], \mathbb{E}_\mu[\theta_1^2])$ are

$$\theta_L = \mathbb{E}_\mu[\theta_1] - (1-\lambda)\sqrt{\frac{\mathbb{E}_\mu[\theta_1^2] - (\mathbb{E}_\mu[\theta_1])^2}{\lambda(1-\lambda)}}, \qquad \theta_H = \mathbb{E}_\mu[\theta_1] + \lambda\sqrt{\frac{\mathbb{E}_\mu[\theta_1^2] - (\mathbb{E}_\mu[\theta_1])^2}{\lambda(1-\lambda)}}$$

for any $\lambda \in (0, 1)$ such that $0 \leqslant \theta_L < \theta_H \leqslant 1$.

Although there is a continuum of two-point mixtures indexed by λ that reproduce the same two-period implications of any given Beta prior, these mixtures yield different third moments $\mathbb{E}_{\mu_\lambda}[\theta^3] = \lambda\theta_L^3 + (1-\lambda)\theta_H^3$, and thus different predictions once a third outcome is observed. We discuss this further in the section on more than two periods.

5.2 Rank of $M(p, q)$ and support of the CIID representation

This section relates the forecasts (p, q) that a prior with support of size r can rationalize to the rank of the matrix $M(p, q)$.

Proposition 3. *If forecast (p, q) admits a CIID representation where μ has an $r \geqslant 2$ point support then* $\operatorname{rank} M(p, q) \leqslant r$.

Proposition 3 is proved by establishing that if (p, q) has a CIID representation with support r, then $\operatorname{cpr}(M(p, q)) \leqslant r$. The result then follows from the general relation cpr $\geqslant$ rank. The next proposition uses the same connection between the size of the support and $\operatorname{cpr}(M(p, q))$, but now paired with the general inequality $\operatorname{cpr} \leqslant \operatorname{rank}(\operatorname{rank}+1)/2 - 1$.

[8] Then $p_1 = \mathbb{E}_\mu[\theta_1]$, $q_1^{(1)} = \mathbb{E}_\mu[\theta_1^2]/\mathbb{E}_\mu[\theta_1]$, $q_1^{(2)} = \mathbb{E}_\mu[\theta_1(1-\theta_1)]/\mathbb{E}_\mu[1-\theta_1]$ and $\mathbb{E}_{\mu_\lambda}[\theta_1] = \lambda\theta_L + (1-\lambda)\theta_H, \mathbb{E}_{\mu_\lambda}[\theta_1^2] = \lambda\theta_L^2 + (1-\lambda)\theta_H^2$.

Proposition 4. *Let forecast (p, q) be such that $M(p, q)$ is completely positive. If M has rank l, then (p, q) admits a CIID representation where μ has at most* $\max\{l(l+1)/2 - 1, 1\}$ *point support.*

Propositions 3 and 4 immediately imply this corollary.

Corollary 5. *For every nondogmatic forecast (p, q) the following statements are equivalent:*

1. *(p, q) admits a CIID representation where μ has a binary support;*
2. *$M(p, q)$ is completely positive and has rank 2.*

Example 3 (A CIID model with support > rank M)**.** *Suppose $Y = \{1, 2\}$, and that forecast (p, q) admits a CIID representation μ with a three point support: $\mu(\theta_1 = \frac{1}{3}) = \frac{1}{4}, \mu(\theta_1 = \frac{1}{2}) = \frac{1}{2}, \mu(\theta_1 = \frac{2}{3}) = \frac{1}{4}$, so that $p_1 = \mathbb{E}_\mu[\theta_1] = \frac{1}{2}$. Then $\mathbb{E}_\mu[\theta_1^2] = \frac{19}{72}$, $\mathbb{E}_\mu[\theta_1(1-\theta_1)] = 17/72$, $q_1^{(2)} = \frac{\mathbb{E}[\theta(1-\theta)]}{\mathbb{E}[1-\theta]} = \frac{17}{36}$, and $q_1^{(1)} = \frac{19}{36}$. Thus the implied second-period conditional forecasts satisfy reinforcement, and*

$$M(p,q) = \begin{pmatrix} p_1 q_1^{(1)} & (1-p_1)q_1^{(2)} \\ p_1(1-q_1^{(1)}) & (1-p_1)(1-q_1^{(2)}) \end{pmatrix} = \begin{pmatrix} \frac{19}{72} & \frac{17}{72} \\ \frac{17}{72} & \frac{19}{72} \end{pmatrix}$$

is symmetric, entrywise nonnegative, and has $\operatorname{rank} M(p, q) = 2 < \operatorname{supp} \mu$.

6 Applications

6.1 Belief in Persistence or Reversal

We next apply our characterization of CIID models to study whether two-period forecasts can distinguish CIID learning from perceived persistence or reversal. We study a simple form of persistence: After observing an outcome, the agent shifts a common fraction of their forecast toward that outcome occurring again. This specification isolates a tendency to predict repetition while leaving the relative probabilities of other outcomes unchanged and allowing an unrestricted CIID baseline. This provides a parsimonious benchmark for studying whether two-period forecasts can distinguish perceived persistence from learning about a fixed distribution.

Definition 7. Forecast (p, q) has a *proportional-persistence representation* if there exists an $\alpha \in (0, 1)$ such that $(p, \hat{q})$ has a CIID representation where

$$q^{(i)} = \alpha e_i + (1 - \alpha)\hat{q}^{(i)} \quad \forall i \in Y, \tag{15}$$

and e_i is the unit vector corresponding to a point mass on the i-th outcome.

An agent whose forecasts have a proportional-persistence representation, while in reality facing an i.i.d. environment, displays what has been called the *hot-hand fallacy* in the line of work pioneered by Gilovich, Vallone, and Tversky [1985].

Proposition 5. *If forecast (p, q) has a proportional-persistence representation then (p, q) has a CIID representation.*

Thus, in two periods, belief in proportional persistence is not distinguishable from a CIID model. It is easy to see that the converse need not be true: If $n = 2$ and $p = q^{(1)} = q^{(2)} = (1/2, 1/2)$, then (p, q) admits a CIID representation with a dogmatic rationalizing belief $\mu = \delta_{(1/2,1/2)}$. However, any proportional-persistence representation (15) would require the associated CIID representation to have $\hat{q}_1^{(1)} < p_1$, a violation of reinforcement, which is not possible by Lemma 1.

Definition 8. Forecast (p, q) has a *proportional-reversal representation* if there exists an $\alpha \in (0, 1)$ such that $(p, \hat{q})$ has a CIID representation where

$$q^{(i)} = \frac{\hat{q}^{(i)}}{(1 - \alpha)} - \frac{\alpha}{(1 - \alpha)} e_i. \tag{16}$$

In words, (p, q) has a proportional-reversal representation if it can be derived from a CIID model that is modified so that after each outcome i, all outcomes $j \neq i$ receive a multiplicative boost of $1/(1 - \alpha)$ to their probability, with the probability of outcome i decreasing accordingly. Having a proportional-reversal representation when facing an i.i.d. environment corresponds to the *gambler's fallacy* as in Tversky and Kahneman, 1971.

Unlike the proportional-persistence representation, some proportional-reversal representations can be spotted from the agent's forecast.[9] However, any two-period

[9] To see this, let $(p, \hat{q})$ be the CIID representation derived from the dogmatic belief $\mu = \delta_{(1/2,1/2)}$ and set $q^{(i)} = 1/2e_{-i} + 1/2\hat{q}$. Then (p, q) has a proportional-reversal representation, but it violates

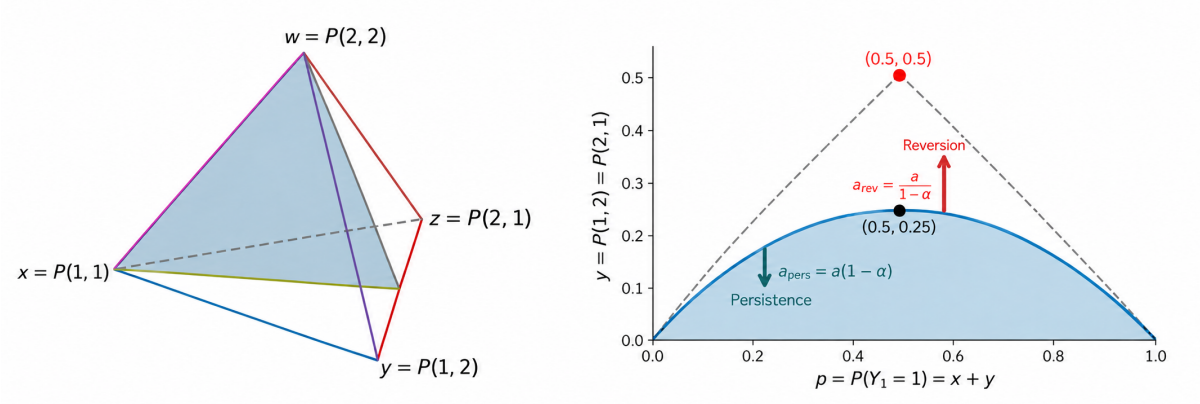


Figure 2: Left: Simplex over the four possible outcome sequences. Shaded region represents the restriction imposed by Symmetry. Right: restrictions on the symmetric probabilities. Persistence moves the probabilities down; reversal moves them up.

forecasts consistent with CIID are also consistent with proportional reversal.

Proposition 6. *If (p, q) has a CIID representation, then (p, q) has a proportional-reversal representation.*

We conclude by considering an agent who believes outcomes are conditionally i.i.d. but observes them with noise. The agent forecasts the next true outcome conditional on a garbled observation of the first. These forecasts can violate symmetry even with binary outcomes and arbitrarily small observation noise: $p_i q_j^{(\tilde{i})} \neq p_j q_i^{(\tilde{j})}$, where $\tilde{i}$ denotes the observed signal corresponding to outcome i. Thus, our test can detect a belief in noisy observations of a CIID process.[10]

6.2 Existence of a Symmetric Common Prior over Signals

We now show how our results can be used to verify the existence of a symmetric common prior over signals from the players' interim beliefs and their ex-ante beliefs about their own type. Formally, consider a game of incomplete information with two

reinforcement, so it does not admit a CIID representation by Lemma 1.

[10]For a concrete example, suppose the agent believes that the true outcome is 1 with probability 1/5 or 3/4 with equal probability, and that the garbling matrix is $G = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$. The agent's period-2 forecast is conditioned on the garbled signal rather than the true outcome, and a straightforward computation shows that symmetry fails.

players and state space $\Omega \times S_1 \times S_2$ with $S_1 = S_2 = S$ finite and $|S| = n$. Here, $\omega \in \Omega$ is an underlying fundamental in the game, and s_i corresponds to player i type, as in Dekel, Fudenberg, and Morris [2007].

A *symmetric common prior over signals* is a measure $\pi \in \Delta(\Omega \times S_1 \times S_2)$ such that

$$\sum_{\omega \in \Omega} \pi(\omega, s_1, s_2) = \int_{\Delta(S)} \theta(s_1)\theta(s_2)\, \mathrm{d}\mu(\theta) \quad \text{for some } \mu \in \Delta(\Delta(S)).$$

With this belief structure, players are certain that their types are i.i.d. draws from the same distribution and share the same belief about that distribution. One example is the case in which the players' signals jointly and completely reveal the state: if knowing (s_1, s_2) determines ω with certainty, then a symmetric common prior over signals pins down all uncertainty about ω as well, so a symmetric common prior over signals implies a symmetric common prior over the entire state space $\Omega \times S_1 \times S_2$. Another example arises in common-value auctions, where ω denotes the value of the good and agents' signals are i.i.d. conditional on ω, with distribution $\hat{\pi}(\cdot \mid \omega)$. In this case, a common prior over signals exists and is given by $\mu(C) = \pi\big(\{\omega : \hat{\pi}(\cdot \mid \omega) \in C\}\big)$ for all $C \in \Delta(S)$.[11] Finally, symmetric common priors over signals are shown to be priors that, in beauty contests, maintain tractability outside Gaussian information structure [Volij, 2026].

Let $\pi_i^{-i}(\cdot|s_i) \in \Delta(S_{-i})$ denote the interim belief of a player i of type s_i about the opponent type. Also, let $\pi_i^{\text{eo}} \in \Delta(S_i)$ denote the ex-ante belief about their own private signal. Our results help us link the existence of a symmetric common prior to the matrices $M_i(\pi) \in [0,1]^{S_{-i} \times S_i}$, $i \in \{1,2\}$ given by $m_i(\pi)_{s_{-i}, s_i} = \pi_i^{\text{eo}}(s_i)\pi_i^{-i}(s_{-i}|s_i)$.

Corollary 6. *Given π_i^{-i} and π_i^{eo} there exists a symmetric common prior over signals consistent with those beliefs if and only if $M_1(\pi) = M_2(\pi)$ and is completely positive.*

The result follows from the fact that, by Theorem 2, the complete positivity of $M_i(\pi)$ is equivalent to the existence of a prior over signals μ_i that is CIID. Then the equality $M_1(\pi) = M_2(\pi)$ guarantees that this prior is common to both players.

[11] Although this structure is typical in common-value auctions, the only thing that is needed is that the values of the good for each agent i can be expressed as functions of θ.

6.3 Grether updating and CIID rationalizability

Consider a Grether model with finite state space $\mathcal{S} = \{1, \ldots, K\}$. The agent assigns prior probability $\lambda_s > 0$ to state s, where $\sum_{s\in\mathcal{S}} \lambda_s = 1$. Conditional on state s, outcomes are i.i.d. with distribution $\theta^s \in \Delta(Y)$. Assume that $\theta^s_i > 0$ for every s and i. Under Grether updating, after observing outcome i, posterior odds between states s and r satisfy

$$\frac{w_s(i)}{w_r(i)} = \left(\frac{\theta^s_i}{\theta^r_i}\right)^c \left(\frac{\lambda_s}{\lambda_r}\right)^d, \qquad c, d > 0.$$

Equivalently, $w_s(i) = \lambda_s^d(\theta_i^s)^c / \sum_{r\in\mathcal{S}} \lambda_r^d(\theta_i^r)^c$. The initial forecast and the forecast after observing i are, respectively, $p_i = \sum_{s\in\mathcal{S}} \lambda_s \theta^s_i$ and $q_j^{(i)} = \sum_{s\in\mathcal{S}} w_s(i)\theta^s_j$. Thus the forecast matrix M^G, is given by

$$M^G_{ji} = \sum_{t\in\mathcal{S}} \lambda_t \theta^t_i \frac{\sum_{s\in\mathcal{S}} \lambda_s^d(\theta_i^s)^c \theta^s_j}{\sum_{r\in\mathcal{S}} \lambda_r^d(\theta_i^r)^c}.$$

Theorem 2 immediately gives the following corollary:

Corollary 7. *The forecasts generated by a finite-state Grether model are CIID-rationalizable if and only if M^G is completely positive.*

When the prior has support on only two states, symmetry implies complete positivity, and there is a simple test for when symmetry holds. Suppose henceforth that the states are A and B, with prior probabilities $\lambda \in (0, 1)$ and $1 - \lambda$, and that their outcome distributions are $a = (a_i)_{i\in Y}$ and $b = (b_i)_{i\in Y}$, where $a_i, b_i > 0$ for every i. After observing i, the posterior probability of A is

$$w_i = \frac{\lambda^d a_i^c}{\lambda^d a_i^c + (1-\lambda)^d b_i^c}.$$

Consequently, $p_i = \lambda a_i + (1 - \lambda)b_i$ and $q_j^{(i)} = w_i a_j + (1 - w_i)b_j$.

Proposition 7. *The forecasts generated by a two-state Grether model are CIID-rationalizable if and only if they satisfy symmetry. If $a \neq b$, these conditions are equivalent to the existence of a scalar $\kappa > 0$ such that $p_i(w_i - \lambda) = \kappa(a_i - b_i)$ for every $i \in Y$. If $a = b$, symmetry and CIID rationalizability always hold.*

With two states and binary outcomes, the symmetry condition reduces to a single

equation: when $a \neq b$ and the prior over states is unbalanced, fixing (λ, a, b, c) selects at most one value of the prior-weight exponent d for which Grether updating is CIID-rationalizable. The explicit formula is given in Online Appendix B.3.

7 CIID Models for More than Two Periods

When the analyst has elicited one-period-ahead forecasts over a longer time horizon T, the CIID representation imposes additional constraints, because longer horizons identify higher-order moments of the latent measure, and imposes tighter necessary conditions. This means fewer forecasts will have CIID representations, and those that do will have fewer.

In the binary case, observing one-step-ahead forecasts up to horizon T identifies the first T moments of the latent Bernoulli parameter θ. We use this to connect to the *Hausdorff truncated moment problem*. This yields an exact and easily checked characterization of CIID models in terms of observable forecasts. At the same time, the truncation point T determines how tightly the prior is pinned down: For interior moment sequences, when T is odd, the minimal-support rationalization is unique, whereas with T even there is a one-dimensional family of distinct minimal-support priors that generate the same finite sequence of forecasts.

Similarly, in the multinomial case, forecasts over more than two periods deliver information about higher-order joint probabilities of outcomes and thus about higher-order moments of the latent θ. The relevant objects are the sequence of moment tensors $\{M^{(k)}\}_{k \leqslant T}$ associated with the joint distributions of $(Y_1, \ldots, Y_k)$ and with the multinomial count moments. Theorem 6 shows that a forecast system up to horizon T has a CIID representation if and only if these tensors are completely positive and satisfy pairwise exchangeability and reinforcement; the matrix test based on complete positivity of $M(p, q)$ is the special case $T = 2$.

Multiple periods let us distinguish between misspecified models that are observationally indistinguishable in two-period data. For example, in the two-period setting, we showed that an agent who believes in spurious persistence (a hot-hand bias implemented by shifting probability weight toward the most recent outcome) cannot be distinguished from a CIID learner: this perturbation preserves the complete pos-

itivity of $M(p,q)$. When forecasts are observed over longer histories, this invariance breaks down, as the finite-horizon CIID conditions of Theorem 6 converge to exchangeability: a forecast system satisfying them for every finite horizon generates an exchangeable measure on infinite sequences, recovering the classical characterizations of de Finetti [1937] and Hewitt and Savage [1955]. Additional periods impose further restrictions on forecasts and narrow the set of rationalizing priors.

7.1 Necessary Conditions for $T > 2$

As in the two-period case, we begin with some intuitive necessary conditions that any CIID representation must satisfy. Fix a horizon $T > 2$. For $t \in \{0, ..., T-1\}$, a history of length t is $h_t = (y_1, \ldots, y_t) \in Y^t$, with $\varnothing$ denoting the empty history ($t = 0$). For any history $h_t \in Y^t$ with $0 \leqslant t < T$, let $q^{(h_t)} \in \Delta(Y)$ denote the period-t forecast of the period-$(t+1)$ outcome, and define $\mathbf{q} = \{q^{(h_t)}\}_{t\in\{0,\ldots,T-1\}, h_t \in Y^t}$. Define the probabilities induced by the forecasts recursively by $P_{\mathbf{q}}(\varnothing) = 1$ and $P_{\mathbf{q}}(h_t, i) = P_{\mathbf{q}}(h_t) q_i^{(h_t)}$.

Definition 9. Forecast $\mathbf{q}$ has a *CIID representation* if there exists a probability measure $\mu \in \Delta(\Delta(Y))$ such that:

(i) For all $i \in Y$:
$$q_i^{(\varnothing)} = \int_{\Delta(Y)} \theta_i \, \mathrm{d}\mu(\theta).$$

(ii) For all $t \in \{1, ..., T-1\}$, $h_t = (y_1, ..., y_t) \in Y^t$ such that $P_{\mathbf{q}}(h_t) > 0$, the posterior measure $\mu(\cdot|h_t)$ is defined by Bayes' rule: for any Borel set $A \subseteq \Delta(Y)$,
$$\mu(A|h_t) = \frac{\int_{\theta\in A} \prod_{s=1}^{t} \theta_{y_s} \, \mathrm{d}\mu(\theta)}{\int_{\theta\in\Delta(Y)} \prod_{s=1}^{t} \theta_{y_s} \, \mathrm{d}\mu(\theta)}$$
and for all $i \in Y$ the conditional forecast satisfies $q_i^{(h_t)} = \int_{\Delta(Y)} \theta_i \, \mathrm{d}\mu(\theta|h_t)$.

(iii) If $P_{\mathbf{q}}(h_t) = 0$ the value of $q^{(h_t)}$ is unrestricted.

Point (iii) leaves the belief after zero probability events completely unrestricted.[12]

[12] One could alternatively require that beliefs are updated using Bayes rule whenever possible, as in perfect extended-Bayesian equilibrium (Fudenberg and Tirole, 1991); the results would be essentially unchanged but slightly more cumbersome to state.

Perhaps the most basic necessary condition for CIID forecasts is that depends only on the count; this was vacuously satisfied in the two-period setting. Define ν by $\nu_i(h_t) = \sum_{s=1}^{t} \mathbf{1}\{y_s = i\}, i \in Y$, for all $t \in \{0, ..., T-1\}$ and $h_t \in Y^t$. Note that when outcome i is observed $\nu((h_t, i)) = \nu(h_t) + e_i$, where e_i is the i-th unit vector. Let $\mathcal{N}(\mathbf{q})$ be the set of count vectors $\nu(h_t)$ of histories h_t with $t < T$ and $P_{\mathbf{q}}(h_t) > 0$.

Definition 10 (Count sufficiency)**.** Forecast $\mathbf{q}$ satisfies *count sufficiency* if for any two histories h_t, h_t' with $\nu(h_t) = \nu(h_t') \in \mathcal{N}(\mathbf{q})$,

$$q^{(h_t)} = q^{(h_t')} =: q^{(\nu)}. \tag{17}$$

We will assume throughout that count sufficiency is satisfied, and use it to simplify the statements of other conditions. We also restrict attention to histories with ex-ante positive probability.

To extend proportional persistence beyond two periods, suppose $\widehat{\mathbf{q}}$ is a CIID forecast system and, after each nonempty history $h_t = (y_1, \ldots, y_t)$, set $q^{(h_t)} = \alpha e_{y_t} + (1-\alpha)\hat{q}^{(h_t)}$, $\alpha \in (0, 1)$, with the same initial forecast and a common α. At two positive-probability histories with identical counts, the CIID baseline forecasts coincide. If one history ends in outcome i and the other does not, their forecasts of i therefore differ by exactly α, violating count sufficiency. Thus proportional persistence can be distinguished from CIID learning using one-step-ahead forecasts.

In addition, CIID representations imply generalizations of the symmetry and reinforcement conditions. The next condition generalizes symmetry.

Definition 11. Forecast $\mathbf{q}$ satisfies *pairwise exchangeability* if, for every $t \in \{0, \ldots, T-2\}$, every history $h_t \in Y^t$ with $P_{\mathbf{q}}(h_t) > 0$, and every $i, j \in Y$,

$$q_i^{(h_t)}\, q_j^{(h_t, i)} = q_j^{(h_t)}\, q_i^{(h_t, j)}.$$

Because any permutation of $\{1, \ldots, T\}$ can be written as a composition of adjacent transpositions, under count invariance pairwise exchangeability up to period T implies invariance under all permutations of length T, which is the definition of exchangeability.

Definition 12. Forecast $\mathbf{q}$ satisfies *reinforcement* if for every $\nu \in \mathcal{N}(\mathbf{q})$ and $i \in Y$

such that $\nu + e_i \in \mathcal{N}(\mathbf{q})$,

$$q_i^{(\nu+e_i)} \;\geqslant\; q_i^{(\nu)}. \tag{18}$$

Lemma 3. *If forecast* $\mathbf{q}$ *has a CIID representation, it satisfies pairwise exchangeability and reinforcement.*

Definition 13 (Martingale property)**.** Forecast $\mathbf{q}$ satisfies the *martingale property* if for every $\nu \in \mathcal{N}(\mathbf{q})$ such that $\nu + e_j \in \mathcal{N}(\mathbf{q})$ for all $j \in Y$,

$$q^{(\nu)} \;=\; \sum_{j \in Y} q_j^{(\nu)}\, q^{(\nu+e_j)}. \tag{19}$$

That is, the agent's current one–step–ahead forecast equals the expectation of their next-period forecast with respect to the agent's current beliefs.

Lemma 4. *If forecast* $\mathbf{q}$ *satisfies pairwise exchangeability then it satisfies the martingale property.*

Note that this lemma is not a statement about a probability measure but only about the reported forecasts.

Moments and Forecasts As in the case $T = 2$, the sufficient conditions for CIID developed below apply to the moments associated with the forecasts.

Definition 14 (Implied Moments)**.** The *implied moments* of forecast $\mathbf{q}$ are *well-defined* if, for every $t \in \{0, \ldots, T\}$, any two histories $h_t, h_t' \in Y^t$ with the same count vector satisfy $P_{\mathbf{q}}(h_t) = P_{\mathbf{q}}(h_t')$. In that case, for every count vector ν with $\|\nu\|_1 \leqslant T$, define $m_\nu := P_{\mathbf{q}}(h_t)$, where h_t is any history with count vector ν. In particular, $m_{\mathbf{0}} = 1$ and, for $\|\nu\|_1 < T$, $m_{\nu+e_i} = m_\nu\, {}_i^{(\nu)}$ if $m_\nu > 0$ and $m_{\nu+e_i} = m_\nu\, p_i^{(\nu)}$

Under the maintained assumption of count sufficiency, pairwise exchangeability guarantees that the implied moments are well-defined, as the next lemma shows.

Lemma 5.

1. *If forecast* $\mathbf{q}$ *is pairwise exchangeable, its implied moments are well-defined.*

2. *If the implied moments of forecast* $\mathbf{q}$ *are well-defined and there is a measure* $\mu \in \Delta(\Delta(Y))$ *that generates them, then* μ *is a CIID representation of* $\mathbf{q}$.

7.2 Binary Outcomes and More than Two Periods

Let $Y = \{1, 2\}$. In this case, if there is a CIID representation we can work with a one-dimensional latent state θ, and count vectors take the form $(j, t-j)$ where ν has j 1's and $t-j$ 2's.

In this case there is an easily checked characterization of when such a measure exists. H To state the sufficient condition, define the *Hankel matrices* $\underline{H}$ and $\overline{H}$ as follows: If T is even,

$$\underline{h}_{i,j} = m_{(i+j,0)} \qquad \forall i, j \in \{0, ..., T/2\},$$

$$\overline{h}_{i,j} = -\Delta m_{(i+j+1,0)} \qquad \forall i, j \in \{0, ..., T/2 - 1\}.$$

If T is odd,

$$\underline{h}_{i,j} = m_{(i+j+1,0)} \qquad \forall i, j \in \{0, ..., (T-1)/2\}$$

$$\overline{h}_{i,j} = -\Delta m_{(i+j,0)} \qquad \forall i, j \in \{0, ..., (T-1)/2\},$$

where $\Delta m_{(r,0)} := m_{(r+1,0)} - m_{(r,0)}$.

Positive semidefiniteness of these matrices, paired with pairwise exchangeability, characterizes the CIID representation. The first condition guarantees a measure matching the pure moments (see, e.g., Schmüdgen, 2017); pairwise exchangeability guarantees that the match extends to mixed moments.

Theorem 5. *Let $n = 2$ and $T \geqslant 2$. Forecast $\mathbf{q}$ has a CIID representation if and only if it satisfies pairwise exchangeability and its Hankel matrices are PSD.*

Minimal-support representations. The horizon also determines how tightly the data identify the simplest CIID rationalization. If the Hankel matrices are positive definite, then when $T = 2k - 1$ there is a unique CIID representation with the minimal support size, namely k atoms. When $T = 2k$, the minimal support size is $k+1$, and there is a one-parameter family of such representations. Thus, for example, three periods uniquely identify a two-point prior, whereas four periods are consistent with a continuum of three-point priors. The proof follows from standard results on the truncated Hausdorff moment problem; see Online Appendix B.5.

7.2.1 Law of Small Numbers

Rabin [2002]'s model of the law of small numbers provides a useful application of the multi-period restrictions. The paper assumes that outcomes are binary, $Y = \{1, 2\}$, that conditional on an unknown $\theta \in \Delta(Y)$, the agent believes that observations are drawn without replacement from a finite urn of size N whose composition corresponds to θ, and that their prior assigns positive probability to an urn containing at least two balls of each outcome. For sequences of more than two observations, the urn is renewed after every two draws. Thus, conditional on θ, observing outcome i makes outcome i less likely on the second draw, generating the gambler's fallacy.

When θ is unknown, however, there is a countervailing force: repeated observations of outcome i lead the agent to infer that θ_i is higher. This inference can be strong enough that the agent's forecasts satisfy reinforcement despite the perceived negative dependence conditional on θ, so Rabin's forecasts can pass the two-period CIID test. The next result shows that this observational equivalence disappears with a third period: Rabin's model assigns different probabilities to sequences with the same outcome counts, violating pairwise exchangeability.

Proposition 8. *Consider Rabin [2002]'s binary law-of-small-numbers model with finite urn size N. The forecasts through period 2 admit a CIID representation if and only if $q_1^{(1,0)} \geqslant q_1^{(0,0)}$. If p is strictly positive, the forecasts through period 3 do not admit a CIID representation. In particular, they violate pairwise exchangeability:*

$$q_1^{(1,0)} q_2^{(2,0)} < q_2^{(1,0)} q_1^{(1,1)}.$$

7.2.2 Sharp Predictions for the Third-Period Forecast

Proposition 9. *Suppose $n = 2$ and that the forecasts through period 2 admit a CIID representation. Suppose also that $q_1^{(0,0)} \in (0, 1)$ and $q_1^{(1,0)} < 1$. A collection of period-3 forecasts extends the two-period forecasts to a CIID representation if and only if, for some $x \in \left[q_1^{(1,0)}, 1 - \frac{q_1^{(0,0)}(1-q_1^{(1,0)})^2}{(1-q_1^{(0,0)})q_1^{(1,0)}}\right]$, we have $q_1^{(2,0)} = x$, $q_1^{(1,1)} = \frac{q_1^{(1,0)}(1-x)}{1-q_1^{(1,0)}}$, and $q_1^{(0,2)} = \frac{q_1^{(0,0)}(1 - 2q_1^{(1,0)} + q_1^{(1,0)}x)}{1 - 2q_1^{(0,0)} + q_1^{(0,0)}q_1^{(1,0)}}$. Every value of x in this interval is attainable by a CIID representation consistent with the original two-period forecasts.*

The width of the identified interval is $\frac{(1-a)(a-p)}{(1-p)a}$. Thus, except at the boundaries, two-period forecasts leave a nondegenerate set of possible third-period forecasts. The interval collapses when $a = p$, in which case the agent is certain about the Bernoulli parameter, and when $a = 1$, in which case observing outcome 1 reveals that its latent probability is one. The proof of this result is in Online Appendix Section B.6

7.3 Many Outcomes and Many Periods

7.3.1 Tensor characterization

For binary outcomes, the key objects were the power moments $m_r = \mathbb{E}[\theta^r]$. For general outcomes, the observable counterparts are order-k sequence tensors defined from joint data probabilities. Under a CIID representation, these coincide with the k-th moment tensors of the latent θ. Our main multi-period result shows that CIID rationalizability is equivalent to each of these tensors being simplex completely positive, together with a natural consistency condition implied by the law of total probability.

An agent's forecasts are the one-step-ahead conditional probabilities. If these forecasts depend only on the outcome counts, as required by a CIID model, we can express them directly in terms of the count moments. The key object for the multi-period characterization is the collection of order-k moment tensors, which generalize the second-moment matrix $M(p, q)$ from the two-period case.

Definition 15 (Completely Positive Tensors and Simplex Completely Positive Tensors)**.** An order-k tensor S is *completely positive (CP)* if there exist $r \in \mathbb{N}$ and nonnegative vectors $(b^{(\ell)})_{\ell=1}^r \in (\mathbb{R}^Y_{\geqslant 0})^r$ such that

$$S_{i_1 \cdots i_k} \;=\; \sum_{\ell=1}^{r} \prod_{j=1}^{k} b_{i_j}^{(\ell)}.$$

It is *simplex completely positive (SCP)* if the vectors can be chosen in the simplex: there exist $r \in \mathbb{N}$, weights $(\gamma_\ell)_{\ell=1}^r \in \mathbb{R}^r_+$ with $\sum_{\ell=1}^r \gamma_\ell = 1$, and vectors $(\pi^{(\ell)})_{\ell=1}^r \in \Delta(Y)^r$ such that

$$S_{i_1 \cdots i_k} \;=\; \sum_{\ell=1}^{r} \gamma_\ell \prod_{j=1}^{k} \pi_{i_j}^{(\ell)}.$$

The simplex restriction in the definition of SCP tensors reflects the fact that the latent vectors correspond to probability measures rather than arbitrary nonnegative vectors. In general, simplex complete positivity is a stronger condition than complete positivity; Lemma B.4 shows that for count moment tensors the two conditions are equivalent.

Definition 16 (Order-k Count Moment Tensor)**.** For each $k \in \{1, \dots, T\}$, define the *order-k count moment tensor* $M^{(k)}(\mathbf{q})$ by

$$M^{(k)}_{i_1 \cdots i_k}(\mathbf{q}) := m_{\nu(i_1,\dots,i_k)}(\mathbf{q}) \qquad \text{for all count vectors } \nu \text{ with } ||\nu||_1 = k.$$

Example 4. *If $|Y| = 3$ and $T = 3$, the order-3 count moment tensor can be viewed as a cube formed by stacking three 3×3 matrices, where each matrix varies the first and second period outcomes holding the 3rd outcome fixed If μ is Dirichlet $(1, 1, 1)$, then $M^{(3)}_\nu$ consists of the 3 matrices*

$$M^{(3)}_{\cdot\cdot 1} = \begin{bmatrix} \frac{1}{10} & \frac{1}{30} & \frac{1}{30} \\ \frac{1}{30} & \frac{1}{30} & \frac{1}{60} \\ \frac{1}{30} & \frac{1}{60} & \frac{1}{30} \end{bmatrix} \quad M^{(3)}_{\cdot\cdot 2} = \begin{bmatrix} \frac{1}{30} & \frac{1}{30} & \frac{1}{60} \\ \frac{1}{30} & \frac{1}{10} & \frac{1}{30} \\ \frac{1}{60} & \frac{1}{30} & \frac{1}{30} \end{bmatrix} \quad M^{(3)}_{\cdot\cdot 3} = \begin{bmatrix} \frac{1}{30} & \frac{1}{60} & \frac{1}{30} \\ \frac{1}{60} & \frac{1}{30} & \frac{1}{30} \\ \frac{1}{30} & \frac{1}{30} & \frac{1}{10}. \end{bmatrix}$$

The main result of this section generalizes the CIID characterization of Theorem 2 to more than two periods. The characterization requires that the associated moment tensor is simplex completely positive.

Theorem 6. *Fix a horizon $T \geqslant 2$. A count-sufficient forecast $\mathbf{q}$ admits a CIID representation if and only if the following conditions hold:*

1. Pairwise exchangeability.
2. Complete positivity: *The order-T count moment tensor $M^{(T)}(\mathbf{q})$ is completely positive.*

The proof of sufficiency proceeds as follows. First, Lemma B.4 shows that if a moment tensor is completely positive, it is simplex completely positive, and Lemma 5 shows that pairwise exchangeability implies the implied moments are well-defined and and Lemma 5 shows that pairwise exchangeability makes the implied moments well-defined and satisfy $\sum_{i \in Y} m_{\nu + e_i} = m_\nu$. Lemma *B.1* uses these consistency identities

to show that the same atoms and weights generate every lower-order tensor. The resulting discrete prior therefore matches all implied moments and, by Lemma 5, rationalizes the forecasts.

Next, Lemma B.1 uses the martingale property of the forecasts to show that all lower-order tensors $M^{(k)}$, $k < T$, are simplex completely positive and can all be written as a product of probability vectors with the weights. Carathéodory's theorem and the martingale identities imply that the same weights and atoms generate all lower–order tensors.

As T increases, the additional forecasts impose further moment restrictions and narrow the set of rationalizing priors. If a forecast system admits a CIID representation at every finite horizon, then a single prior rationalizes the entire infinite sequence, whose induced distribution is exchangeable.

8 Discussion

This paper characterizes when a sequence of one-step-ahead forecasts is consistent with a CIID model. For the two-period, binary-outcome case, the conditions are simple and intuitive: symmetry and reinforcement. For more outcomes, these conditions are necessary but not sufficient; the key object becomes the second-moment matrix $M(p, q)$, which must be completely positive. For more periods, the entire hierarchy of moment tensors must be simplex completely positive.

Our results provide a clear, operational way to test whether observed forecasting behavior can be explained by a classic model of learning about a stable, unknown environment. For example, forecasts where outcome i is most reinforced after outcome j and vice versa, as in Example 1, can be immediately flagged as non-Bayesian in this sense. More generally, any failure of complete positivity (or, in the $n \leqslant 4$ case, positive semidefiniteness) shows that the agent's updating rule is inconsistent with any CIID representation.

Our proportional-persistence result highlights a limit of two-period forecast data. Any forecast with a proportional-persistence representation also admits a CIID representation, because the perturbation preserves complete positivity of the second-moment forecast matrix. This observational equivalence need not extend to forecasts

after longer histories. Applying the same perturbation to the most recent outcome after each history can violate count sufficiency, assigning different forecasts to histories with identical outcome counts.

References

Aldous, D. J., I. A. Ibragimov, and J. Jacod (2006). *Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983*. Vol. 1117. Springer.

Aliprantis, C. and K. Border (2013). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer-Verlag.

Berman, A. and N. Shaked-Monderer (2003). *Completely positive matrices*. World Scientific.

Bohren, J. A. and D. N. Hauser (2026). "The behavioral foundations of model misspecification: A decomposition." *Working paper*.

Buscemi, F. (2026). "Witnessing Quantum Bayesian Inference beyond Classical Learning." *arXiv preprint arXiv:2610.09293*.

Catonini, E. and G. Lanzani (2026). "A Dutch-book trap for misspecification." *mimeo*.

Chen, X., L. P. Hansen, and P. G. Hansen (2020). "Robust identification of investor beliefs." *Proceedings of the National Academy of Sciences*, 117, 33130–33140.

Cheney, W. (2013). *Analysis for Applied Mathematics*. Springer Science & Business Media.

Dekel, E., D. Fudenberg, and S. Morris (2007). "Interim correlated rationalizability." *Theoretical Economics*, 2, 15–40.

Diaconis, P. (1977). "Finite forms of de Finetti's theorem on exchangeability." *Synthese*, 36, 271–281.

Dominitz, J. and C. F. Manski (2011). "Measuring and interpreting expectations of equity returns." *Journal of applied econometrics*, 26, 352–370.

Epstein, L. G. and M. Le Breton (1993). "Dynamically consistent beliefs must be Bayesian." *Journal of Economic theory*, 61, 1–22.

Finetti, B. de (1937). "La prévision: ses lois logiques, ses sources subjectives." *Annales de l'institut Henri Poincaré*. Vol. 7, 1–68.

Fudenberg, D. and J. Tirole (1991). "Perfect Bayesian equilibrium and sequential equilibrium." *journal of Economic Theory*, 53, 236–260.

Ghirardato, P. (2002). "Revisiting Savage in a conditional world." *Economic theory*, 20, 83–92.

Gilovich, T., R. Vallone, and A. Tversky (1985). "The hot hand in basketball: On the misperception of random sequences." *Cognitive psychology*, 17, 295–314.

Golub, B. and S. Morris (2020). "Expectations, networks, and conventions." *arXiv preprint arXiv:2009.13802*.

Green, E. J. and I.-U. Park (1996). "Bayes contingent plans." *Journal of Economic Behavior & Organization*, 31, 225–236.

Greenwood, R. and A. Shleifer (2014). "Expectations of returns and expected returns." *The Review of Financial Studies*, 27, 714–746.

Grether, D. M. (1980). "Bayes rule as a descriptive model: The representativeness heuristic." *The Quarterly journal of economics*, 95, 537–557.

Hewitt, E. and L. J. Savage (1955). "Symmetric measures on Cartesian products." *Transactions of the American Mathematical Society*, 80, 470–501.

Maxfield, J. E. and H. Minc (1962). "On the Matrix Equation $X'X = A$." *Proceedings of the Edinburgh Mathematical Society*, 13, 125–129.

Molavi, P. (2026). "Misspecified Bayesianism." *arXiv preprint arXiv:2109.07007*.

Rabin, M. (2002). "Inference by believers in the law of small numbers." *The Quarterly Journal of Economics*, 117, 775–816.

Sarnoff, K. (2026). "The Structure of Sequential Updating." *Working paper*.

Schmüdgen, K. (2017). *The moment problem*. Springer.

Shmaya, E. and L. Yariv (2016). "Experiments on decisions under uncertainty: A theoretical framework." *American Economic Review*, 106, 1775–1801.

Tversky, A. and D. Kahneman (1971). "Belief in the law of small numbers." *Psychological bulletin*, 76, 105.

Volij, O. (2026). "Beauty Contests: Beyond Gaussian Uncertainty."

Weber, M., F. d'Acunto, Y. Gorodnichenko, and O. Coibion (2022). "The subjective inflation expectations of households and firms: Measurement, determinants, and implications." *Journal of Economic Perspectives*, 36, 157–184.

A Appendix

Proof of Theorem 1. Lemma 1 shows that symmetry and reinforcement are necessary. To show they are sufficient, suppose forecast (p, q) satisfies symmetry and

reinforcement.

Case a Suppose $1 > p_1 > q_1^{(2)} > 0$ and take the Beta parameters from (4). Then

$$\tilde{p}_1 = \frac{\alpha}{\alpha+\beta} = \frac{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}}}{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}} + \frac{q_1^{(2)}(1-p_1)}{p_1 - q_1^{(2)}}} = \frac{p_1 q_1^{(2)}}{p_1 q_1^{(2)} + q_1^{(2)}(1-p_1)} = p_1.$$

The period-2 forecasts satisfy $\tilde{q}_1^{(2)} = \frac{\alpha}{\alpha+\beta+1} = \frac{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}}}{\frac{p_1}{p_1 - q_1^{(2)}}} = q_1^{(2)}$ and $\tilde{q}_1^{(1)} = \frac{\alpha+1}{\alpha+\beta+1} = \frac{\alpha+1}{\alpha}\tilde{q}_1^{(2)}$. Since $\frac{\alpha+1}{\alpha} = \frac{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}} + 1}{\frac{p_1 q_1^{(2)}}{p_1 - q_1^{(2)}}} = \frac{p_1 q_1^{(2)} + p_1 - q_1^{(2)}}{p_1 q_1^{(2)}}$, we obtain $\tilde{q}_1^{(1)} = \frac{p_1 q_1^{(2)} + p_1 - q_1^{(2)}}{p_1} = q_1^{(2)} + 1 - \frac{q_1^{(2)}}{p_1}$. By symmetry $p_1(1-q_1^{(1)}) = (1-p_1)q_1^{(2)}$, hence $q_1^{(1)} = q_1^{(2)} + 1 - \frac{q_1^{(2)}}{p_1} = \tilde{q}_1^{(1)}$, completing the proof for this case.

Case b Suppose $p_1 = q_1^{(2)}$. If $p_1 > 0$, symmetry implies $p_2 = q_2^{(1)}$. If $p_1 = 0$, then $q^{(2)} = (0,1)$, while $q^{(1)}$ is unrestricted because outcome 1 has zero probability. In either case, specifying that μ is a point mass on (p_1, p_2) rationalizes (p, q).

Case c If $p_1 \neq q_1^{(2)}$ and $q_1^{(2)} = 0$, then by symmetry $q_2^{(1)} = 0$. Then specifying $\mu = p_1\delta_{(1,0)} + (1-p_1)\delta_{(0,1)}$ recovers the specified (p, q).

Case d If $p_1 = 1$, then by reinforcement $q_1^{(1)} = 1$ and $q^{(2)}$ is arbitrary, so specifying that μ is a point mass on $\theta = (1,0)$ recovers the specified (p, q).

Since reinforcement requires $p_1 \geqslant q_1^{(2)}$, every forecast satisfying reinforcement falls into exactly one of cases a–d. □

Proof of Lemma 2. *Only if.* Observe that by equations (1) and (3) if $\mu \in \Delta(\Delta(Y))$ is a CIID representation for (p, q) then for all $i, j \in Y$

$$m_{ij}(p,q) = \int_{\Delta(Y)} \theta_j d\mu(\theta) \int_{\Delta(Y)} \theta_i d\mu(\theta|j = \int_{\Delta(Y)} \theta_j d\mu(\theta) \int_{\Delta(Y)} \frac{\theta_i \theta_j}{\int_{\Delta(Y)} \theta_j d\mu(\theta)} d\mu(\theta) = \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta) \tag{20}$$

if $p_j \neq 0$ and $m_{ij}(p,q) = 0 = \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta)$ if $p_j = 0$.

If. Suppose that there exists a $\mu \in \Delta(\Delta(Y))$ such that equation (5) is satisfied. Then $p_j = \sum_{i=1}^n m_{ij}(p,q) = \sum_{i=1}^n \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta) = \int_{\Delta(Y)} \theta_j d\mu(\theta)$, so that (1) holds.

Moreover, if $p_j \neq 0$

$$q_i^{(j)} = \frac{m_{ij}(p,q)}{p_j} = \frac{\int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta)}{\int_{\Delta(Y)} \theta_j d\mu(\theta)} = \int_{\Delta(Y)} \theta_i d\mu(\theta|j),$$

so μ is a CIID model for (p,q) with $\mu(\cdot|j)$ arbitrary for j such that $p_j = 0$. Moreover, $M(p,q)$ is positive semidefinite because it is a mixture of rank-one positive semidefinite matrices, the outer products $\theta\theta^\top$. □

Proof of Claim 1. Suppose that M is completely positive, so there exist k and $\left(\bar{\pi}^{(s)}\right)_{s=1}^k \in (\mathbb{R}_+^n)^k$ such that $m_{ij} = \sum_{s=1}^k \bar{\pi}_i^{(s)} \bar{\pi}_j^{(s)}$ for all $i,j \in \{1,\ldots,n\}$. For each s, let $Z_s := \sum_{l=1}^n \bar{\pi}_l^{(s)}$, $\hat{\pi}_i^{(s)} := \frac{\bar{\pi}_i^{(s)}}{Z_s}$, and $\gamma_s := Z_s^2 \in \mathbb{R}_+$. Then $\hat{\pi}^{(s)} \in \Delta(Y)$ and $m_{ij} = \sum_{s=1}^k \bar{\pi}_i^{(s)} \bar{\pi}_j^{(s)} = \sum_{s=1}^k Z_s^2 \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)} = \sum_{s=1}^k \gamma_s \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)}$. Moreover, $\sum_{i=1}^n m_{ij} = \sum_{s=1}^k \gamma_s \sum_{i=1}^n \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)} = \sum_{s=1}^k \gamma_s \hat{\pi}_j^{(s)}$, so $1 = \sum_{i=1}^n \sum_{j=1}^n m_{ij} = \sum_{s=1}^k \gamma_s \sum_{j=1}^n \hat{\pi}_j^{(s)} = \sum_{s=1}^k \gamma_s$. Thus $\sum_{s=1}^k \gamma_s = 1$ and $m_{ij} = \sum_{s=1}^k \gamma_s \hat{\pi}_i^{(s)} \hat{\pi}_j^{(s)}$ is a simplex completely positive representation, which proves the claim. □

Proof of Theorem 2. (2) ⇒ (1). By Lemma 2, there is $\mu \in \Delta(\Delta(Y))$ with $m_{ij}(p,q) = \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta)$ for every $i,j \in Y$. Since $\Delta(Y)$ is compact and $\pi \mapsto \pi\,\pi^\top$ is continuous, $M(p,q) \in \operatorname{conv}\{\pi\,\pi^\top : \pi \in \Delta(Y)\}$. By Carathéodory's theorem (see Aliprantis and Border, 2013, Theorem 5.32 with dimension $r = \frac{1}{2}n(n+1)$), we can write

$$M(p,q) = \sum_{s=1}^{r+1} \gamma_s\, \pi^{(s)} \pi^{(s)\top}, \text{ where all } \gamma_s > 0, \sum_{s=1}^{r+1} \gamma_s = 1, \text{ and all } \pi^{(s)} \in \Delta(Y) \quad (21)$$

establishing complete positivity.

(1) ⇒ (3). Since $M(p,q)$ is a forecast matrix, its entries sum to one (as shown in the text), so complete positivity implies simplex complete positivity by Claim 1. Given the representation $M(p,q) = \sum_{s=1}^{r} \gamma_s\, \pi^{(s)} \pi^{(s)\top}$, define the discrete measure $\mu = \sum_{s=1}^r \gamma_s\, \delta_{\pi^{(s)}}$. Then $\int_{\Delta(Y)} \pi_i\, \pi_j\, \mathrm{d}\mu(\pi) = \sum_{s=1}^r \gamma_s\, \pi_i^{(s)}\, \pi_j^{(s)} = m_{ij}(p,q)$, establishing that $m_{ij}(p,q) = \int_{\Delta(Y)} \theta_i \theta_j d\mu(\theta)$ for every $i,j \in Y$.

(3) ⇒ (2). Trivial. □

Proof of Corollary 1. For the second part, suppose that (p,q) has a CIID representation. By Theorem 2 (p,q) has a CIID representation with finite support. By Lemma 1 (p,q) satisfies symmetry and reinforcement, so the proof of Theorem 1

case a., and the fact that with binary outcomes a nondogmatic forecast has $p \neq q^{(2)}$, implies there is a CIID model with a Beta distribution. □

Lemma A.1. *Suppose $M(p,q)$ is completely positive with CIID representation μ. For each integer $k \geqslant 1$, draw $\theta^{(1)}, \ldots, \theta^{(k)}$ independently from μ and define*

$$\mu_k = \frac{1}{k}\sum_{s=1}^{k} \delta_{\theta^{(s)}}, \qquad M_k = \frac{1}{k}\sum_{s=1}^{k} \theta^{(s)}\theta^{(s)\top}.$$

Then $\mathbb{E}[\|M_k - M(p,q)\|_F] \leqslant k^{-1/2}$, where the expectation is over these draws. In particular, there is a CIID prior with supported om at most k points whose induced forecast matrix is within Frobenius distance $k^{-1/2}$ of $M(p,q)$.

Proof of Theorem 3. Suppose that $0 < \varepsilon < 1/n$ and that $M(p,q)$ is within Frobenius distance ε from the completely positive cone. Choose a completely positive matrix A such that $\|M(p,q) - A\|_F \leqslant \varepsilon$. Let $a = \sum_{i=1}^n \sum_{j=1}^n A_{ij}$. Since the entries of $M(p,q)$ sum to one, Cauchy–Schwarz gives

$$\begin{aligned} |a-1| &= \left|\sum_{i=1}^{n}\sum_{j=1}^{n} 1 \cdot \big(A_{ij} - M(p,q)_{ij}\big)\right| \leqslant \left(\sum_{i=1}^{n}\sum_{j=1}^{n} 1^2\right)^{1/2} \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\big(A_{ij} - M(p,q)_{ij}\big)^2\right)^{1/2} \\ &= n\,\|A - M(p,q)\|_F \leqslant n\varepsilon. \end{aligned}$$

Consequently, if $\varepsilon < 1/n$, then $a \geqslant 1 - n\varepsilon > 0$, so we can normalize A by dividing by a. The matrix $\bar{A} = A/a$ is completely positive and its entries sum to one, so Claim 1 implies that it is generated by a CIID representation. Since $\bar{A}$ has nonnegative entries summing to one, $\|\bar{A}\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n \bar{A}_{ij}^2 \leqslant \left(\sum_{i=1}^n \sum_{j=1}^n \bar{A}_{ij}\right)^2 = 1$. Taking square roots gives $\|\bar{A}\|_F \leqslant 1$. Moreover, $A = a\bar{A}$, so $\bar{A} - A = (1-a)\bar{A}$. Thus, $\|\bar{A} - A\|_F = |1-a|\,\|\bar{A}\|_F \leqslant |1-a| \leqslant n\varepsilon$, where the equality uses the scaling property of the Frobenius norm and the final inequality is the Cauchy–Schwarz bound established above. Finally, the triangle inequality gives

$$\|\bar{A} - M(p,q)\|_F = \|(\bar{A} - A) + (A - M(p,q))\|_F \leqslant \|\bar{A} - A\|_F + \|A - M(p,q)\|_F \leqslant n\varepsilon + \varepsilon = (n+1)\varepsilon.$$

Applying Lemma A.1 to a CIID representation of $\bar{A}$ and taking $k = \lceil \varepsilon^{-2} \rceil$, there is a measure with at most k atoms whose induced matrix B satisfies $\|B - \bar{A}\|_F \leqslant k^{-1/2} \leqslant \varepsilon$. Therefore, $\|B - M(p,q)\|_F \leqslant \|B - \bar{A}\|_F + \|\bar{A} - M(p,q)\|_F \leqslant (n+2)\varepsilon$. For $\varepsilon \geqslant 1/n$,

instead let $u = (1/n, \ldots, 1/n)$ and take $B = uu^\top$, the matrix generated by the Dirac CIID δ_u. Since $M(p,q)$ has nonnegative entries summing to one, $\|M(p,q)\|_F^2 \leqslant 1$. Every entry of $B = uu^\top$ equals $1/n^2$, so expanding the squares gives

$$\begin{aligned}\|B - M(p,q)\|_F^2 &= \sum_{i=1}^n \sum_{j=1}^n \left(m_{ij}(p,q) - \frac{1}{n^2}\right)^2 = \sum_{i=1}^n \sum_{j=1}^n m_{ij}(p,q)^2 - \frac{2}{n^2}\sum_{i=1}^n \sum_{j=1}^n m_{ij}(p,q) + \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n^4} \\ &= \|M(p,q)\|_F^2 - \frac{2}{n^2} + \frac{1}{n^2} = \|M(p,q)\|_F^2 - \frac{1}{n^2} \leqslant 1.\end{aligned}$$

Thus $\|B - M(p,q)\|_F \leqslant 1 \leqslant n\varepsilon$, and obviously $1 \leqslant \lceil \varepsilon^{-2} \rceil$. In both cases, we have constructed a CIID prior μ' with support at most $\lceil \varepsilon^{-2} \rceil$ whose induced matrix B satisfies $\|B - M(p,q)\|_F \leqslant (n+2)\varepsilon$.

We now recover the forecasts generated by this prior. Define

$$p'_j = \sum_{i=1}^n B_{ij}, \qquad q_i'^{(j)} = \begin{cases} B_{ij}/p'_j, & \text{if } p'_j > 0, \\ q_i^{(j)}, & \text{if } p'_j = 0. \end{cases}$$

Because B is induced by μ', these are its initial and conditional forecasts whenever $p'_j > 0$. When $p'_j = 0$, the CIID definition leaves the conditional forecast unrestricted. Hence (p', q') admits the CIID representation μ' and $M(p', q') = B$, proving (7).

To bound the changes in forecasts, let $E_{ij} = B_{ij} - m_{ij}(p,q)$. Since the column sums of B and $M(p,q)$ are p'_j and p_j, $|p'_j - p_j| = |\sum_{i=1}^n E_{ij}| \leqslant \sum_{i=1}^n |E_{ij}$. For each column, $B_{\cdot j} = p'_j q'^{(j)}$ and $M(p,q)_{\cdot j} = p_j q^{(j)}$, including when either marginal is zero. Thus, $p_j\big(q'^{(j)} - q^{(j)}\big) = E_{\cdot j} + (p_j - p'_j)q'^{(j)}$. Taking ℓ^1 norms and using $\|q'^{(j)}\|_1 = 1$ gives $p_j\|q'^{(j)} - q^{(j)}\|_1 \leqslant \sum_{i=1}^n |E_{ij} + |p_j - p'_j| \leqslant 2d_j$. Cauchy–Schwarz over the n^2 matrix entries yields $\sum_{i=1}^n \sum_{j=1}^n |E_{ij}| \leqslant n\,\|B - M(p,q)\|_F \leqslant n\delta$. Summing the marginal and weighted conditional bounds over j therefore proves (8).

Finally, Cauchy–Schwarz within each column gives $\sum_{i=1}^n |E_{ij} \leqslant \sqrt{n}\,\big(\sum_{i=1}^n E_{ij}^2\big)^{1/2} \leqslant \sqrt{n}\,\|B - M(p,q)\|_F \leqslant \sqrt{n}\,\delta$. If $p_j > 0$ for every j, it follows that

$$\|q'^{(j)} - q^{(j)}\|_1 \leqslant \frac{2\sum_{i=1}^n |E_{ij}}{p_j} \leqslant \frac{2\sqrt{n}\,\delta}{\min_{j\in\{1,\ldots,k\}} p_j} = \frac{2\sqrt{n}\,(n+2)}{\min_{j\in\{1,\ldots,k\}} p_j}\,\varepsilon, \qquad \forall j \in \{1, \ldots, k\}.$$

Taking the maximum over j proves (9). □

Proof of Theorem 4. Any vector x whose entries sum to zero satisfies

$$\|x\|_2 \leqslant \frac{\|x\|_1}{\sqrt{2}}. \tag{22}$$

Indeed, writing $x = x^+ - x^-$, its positive and negative parts have disjoint supports and satisfy $\|x^+\|_1 = \|x^-\|_1 = \|x\|_1/2$. Hence $\|x\|_2^2 = \|x^+\|_2^2 + \|x^-\|_2^2 \leqslant \|x^+\|_1^2 + \|x^-\|_1^2 = \frac{1}{2}\|x\|_1^2$.

By Theorem 2, $M(p', q') \in \mathcal{CP}_n$. Decompose $M(p,q) - M(p',q') = U + V$, $U_{\cdot j} = p_j(q^{(j)} - q'^{(j)})$, and $V_{\cdot j} = (p_j - p'_j)q'^{(j)}$. Since $q^{(j)}$ and $q'^{(j)}$ are probability vectors, (22) gives $\|U\|_F^2 = \sum_{j=1}^n p_j^2 \|q^{(j)} - q'^{(j)}\|_2^2 \leqslant \frac{1}{2}\sum_{j=1}^n p_j^2 \|q^{(j)} - q'^{(j)}\|_1^2$. Also $\|q'^{(j)}\|_2 \leqslant 1$, so $\|V\|_F^2 = \sum_{j=1}^n (p_j - p'_j)^2 \|q'^{(j)}\|_2^2 \leqslant \|p - p'\|_2^2$, so the triangle inequality implies

$$\inf_{A \in \mathcal{CP}_n} \|M(p,q) - A\|_F \leqslant \|M(p,q) - M(p',q')\|_F \leqslant \|p - p'\|_2 + \frac{1}{\sqrt{2}} \left(\sum_{j=1}^n p_j^2 \|q^{(j)} - q'^{(j)}\|_1^2 \right)^{1/2}.$$

Applying (22) to $p - p'$ gives $\|p - p'\|_2 \leqslant \|p - p'\|_1/\sqrt{2}$, and $\left(\sum_j (p_j \|q^{(j)} - q'^{(j)}\|_1)^2\right)^{1/2} \leqslant \sum_j p_j \|q^{(j)} - q'^{(j)}\|_1$. This proves (10). The remaining bounds follow by setting $p' = p$. When $p_j = 0$, the corresponding column and error weight are zero, so the value of $q^{(j)}$ has no effect on these bounds. □

Proof of Proposition 1. By equation (13), a full cycle requires $\mathrm{Cov}_\mu(\theta_{i+1}, \theta_i) \geqslant 0$ for all i along the cycle, with at least one strict inequality.

Case $n = 2$: By Lemma 1, (p, q) satisfies reinforcement. For $i \neq j$, taking complements gives $q_i^{(j)} = 1 - q_j^{(j)} \leqslant 1 - p_j = p_i$. Thus neither full-cycle inequality can be strict, ruling out a full cycle.

Case n=3: From equation (14):

$$\mathrm{Var}_\mu(\theta_1) + \mathrm{Var}_\mu(\theta_2) + \mathrm{Var}_\mu(\theta_3) + 2\left(\mathrm{Cov}_\mu(\theta_1, \theta_2) + \mathrm{Cov}_\mu(\theta_2, \theta_3) + \mathrm{Cov}_\mu(\theta_3, \theta_1)\right) = 0.$$

Since variances are nonnegative, if the sum of the three covariances were strictly positive, the left-hand side would be strictly positive, a contradiction. So at least one of the cycle covariances must be non-positive, and a full cycle cannot occur.

Case n=4: The sum of the four edge covariances is

$$\begin{aligned}&\mathrm{Cov}_\mu(\theta_1, \theta_2) + \mathrm{Cov}_\mu(\theta_2, \theta_3) + \mathrm{Cov}_\mu(\theta_3, \theta_4) + \mathrm{Cov}_\mu(\theta_4, \theta_1)\\ &= \mathrm{Cov}_\mu(\theta_1 + \theta_3, \theta_2 + \theta_4) = \mathrm{Cov}_\mu(\theta_1 + \theta_3, 1 - (\theta_1 + \theta_3)) = -\mathrm{Var}_\mu(\theta_1 + \theta_3) \leqslant 0.\end{aligned}$$

A full cycle is therefore impossible. □

Proof of Proposition 2. Fix $n \geqslant 5$, and choose H, L with $0 < L < H < 1$ satisfying

$$2H + (n-2)L \;=\; 1. \tag{23}$$

For $r \in Y$ define $\theta^{(r)} \in \Delta(Y)$ by $\theta^{(r)}_k = H$ if $k \in \{r,\ r+1 \bmod n\}$, and L otherwise. Let μ be the uniform measure on $\{\theta^{(1)}, \ldots, \theta^{(n)}\}$. By equation (23), $p_j = \mathbb{E}_\mu[\theta_j] = \dfrac{2H + (n-2)L}{n} = \dfrac{1}{n}$ for all $j \in Y$, so the period-1 forecast is uniform.

Fix $i \in Y$. Across the n support points, the pair (θ_i, θ_{i+1}) takes values (H,H) once, (H,L) and (L,H) once each, and (L,L) the remaining $n-3$ times. Hence, $\mathbb{E}_\mu[\theta_i\theta_{i+1}] = \frac{1}{n}\Big(H^2 + 2HL + (n-3)L^2\Big)$ and $\mathbb{E}_\mu[\theta_i] \;=\; \frac{1}{n}$. Thus, by equation (13), $q^{(i)}_{i+1} \;=\; \dfrac{\mathbb{E}_\mu[\theta_i\theta_{i+1}]}{\mathbb{E}_\mu[\theta_i]} \;=\; H^2 + 2HL + (n-3)L^2$. We need $q^{(i)}_{i+1} > p_{i+1} = 1/n$. Define $f_n(L) := H(L)^2 + 2H(L)L + (n-3)L^2$, where $H(L) := \frac{1-(n-2)L}{2}$. A direct calculation gives $f_n(L) \;=\; \frac{1}{4} - \frac{n-4}{2}L + \frac{n(n-4)}{4}L^2$, so $f_n(0) \;=\; \frac{1}{4}$. For $n \geqslant 5$, we have $f_n(0) = 1/4 > 1/n$. By continuity of $f_n(L)$, there exists $\varepsilon > 0$ such that $f_n(L) > 1/n$ for all $L \in (0, \varepsilon)$. Picking such an L and setting $H = \dfrac{1-(n-2)L}{2}$ yields $q^{(i)}_{i+1} > 1/n = p_{i+1}$ for every $i \in Y$, which constitutes a full cycle. □

Proof of Corollary 4. Part (i). By assumption there is a positive diagonal matrix $D = \mathrm{diag}(s_1, \ldots, s_n)$ with $s \in \mathbb{R}^n_{++}$ such that $(DM(p,q)D)_{ii} \;\geqslant\; \sum_{j \neq i}(DM(p,q)D)_{ij}$ for all $i \in Y$. Therefore $A := DM(p,q)D$ is symmetric, nonnegative, and diagonally dominant, hence it is completely positive by Theorem 2.5 in Berman and Shaked-Monderer [2003]. As a consequence, it can be written as $A = \sum_{k=1}^r \alpha_k u^{(k)}(u^{(k)})^\top$ for some $\alpha \in \mathbb{R}^k$, $(u^{(k)})_{k=1}^r \in (\mathbb{R}^n)^k$. Conjugating by D^{-1} yields $M(p,q) \;=\; D^{-1}AD^{-1} \;=\; \sum_{k=1}^r \alpha_k \big(D^{-1}u^{(k)}\big)\big(D^{-1}u^{(k)}\big)^\top$, which is a sum of nonnegative rank-one outer products. Hence $M(p,q)$ is completely positive. With this, (p,q) admits a CIID representation by Theorem 2.

Part (ii). The inequalities $m_{ii}(p,q)s_i \geqslant \sum_{j \neq i} m_{ij}(p,q)s_j$ for all $i \in Y$ are equivalent to $s \geqslant R(p,q)s$. If $\rho(R) < 1$ then by the Neumann series lemma (see, e.g., Cheney, 2013) $s := (I-R)^{-1}\mathbf{1} \;=\; \sum_{k=0}^\infty R^k\mathbf{1}$ is well-defined and strictly positive. Moreover, $(I-R)s = \mathbf{1}$ implies $s - Rs = \mathbf{1} \geqslant 0$, i.e. $s \geqslant Rs$, and rescaling s to be a probability measure shows that (p,q) satisfies scaled diagonal dominance. □

Proof of Proposition 3. Let $(\theta(1), ..., \theta(r))$ be the elements of the support of μ, and define $\phi(i) = \mu(\theta(i))$. Define the vectors $f_i \in \mathbb{R}^r$ by $f_i = \left(\sqrt{\phi(1)}\theta_i(1), ..., \sqrt{\phi(r)}\theta_i(r)\right)^\top$ and let M be defined by $m_{ij} = \langle f_i, f_j \rangle$, so that $m_{ij} = \sum_{k=1}^r (\sqrt{\phi(k)}\theta_i(k))(\sqrt{\phi(k)}\theta_j(k)) = \sum_{k=1}^r \phi(k)\theta_i(k)\theta_j(k)$. Bayes rule in the CIID model with prior μ gives

$$q_j^{(i)} = \sum_{k=1}^r \frac{\phi(k)\theta_i(k)}{\sum_{\kappa=1}^r \phi(\kappa)\theta_i(\kappa)}\theta_j(k) \tag{24}$$

and $m_{ij}(p,q) = p_i q_j^{(i)} = \sum_{k=1}^r \phi(k)\theta_i(k)q_j^{(i)} = \sum_{k=1}^r \phi(k)\theta_i(k)\theta_j(k) = m_{ij}$, where the third equality follows from equation (24). Thus the CP-rank of M is at most r and by Proposition 3.2 in Berman and Shaked-Monderer [2003], $\operatorname{rank} M \leqslant \operatorname{cpr}(M)$. □

Proof of Proposition 4. If $\operatorname{rank} M(p,q) = 1$, complete positivity gives $M(p,q) = xx^\top$ for some $x \geqslant 0$. Since the entries of $M(p,q)$ sum to one, $\sum_{i\in Y} x_i = 1$. Its column sums therefore imply $x = p$, so $M(p,q) = pp^\top$. By Lemma 2, δ_p rationalizes (p,q), establishing the bound for rank one. So in what follows, suppose $l \geqslant 2$. Since $M(p,q)$ is completely positive, $m_{ij}(p,q) = \sum_{l=1}^r x_i(l)x_j(l)$ with $x(l) \in \mathbb{R}^n_+$ for each l. Moreover, by Theorem 3.5 in Berman and Shaked-Monderer [2003], when $\operatorname{rank} M = l \geqslant 2$ we can pick these $(x(l))_{l=1}^r$ such that $r \leqslant l(l+1)/2 - 1$.

Without loss of generality, we may assume that each $x(l)$ is nonzero and that no two vectors $x(k)$ and $x(l)$ are proportional: Any zero vector contributes nothing to the sum and may be omitted, and if $x(k) = \lambda x(l)$ for some $\lambda > 0$, then $x(k)x(k)^\top + x(l)x(l)^\top = (\lambda^2 + 1)x(l)x(l)^\top$, so the two terms can be merged into a single rank-one term without changing $M(p,q)$.

Let $S_k = \sum_{i\in Y} x_i(k)$ for all $k \in \{1, ..., r\}$. Because $1 = \sum_{i,j\in Y} m_{i,j}(p,q) = \sum_{i,j\in Y}\sum_{k=1}^r x_i(k)x_j(k) = \sum_{k=1}^r (\sum_{i\in Y} x_i(k))(\sum_{j\in Y} x_j(k)) = \sum_{k=1}^r S_k^2$, $\sum_{k=1}^r S_k^2 = 1$. Define $\phi(k) = S_k^2$. Since $S_k > 0$ and $\sum_{k=1}^r S_k^2 = 1$, we have $\phi(k) \in (0,1)$. Define $\theta_i(k) = \frac{x_i(k)}{S_k}$ $\forall k \in \{1, ..., r\}$ Since $\sum_{i\in Y} x_i(k) = S_k$, each $\theta(k)$ are probability vectors in Δ^{n-1}. We now check that the CIID model with prior μ supported on

$(\theta(1), ..., \theta(r))$ and with $\mu(\theta(k)) = \phi(k)$ induces (p, q). Note that

$$p_j = \sum_{i \in Y} m_{ij} = \sum_{i \in Y} \sum_{k=1}^{r} x_i(k) x_j(k) = \sum_{k=1}^{r} x_j(k) \sum_{i \in Y} x_i(k) = \sum_{k=1}^{r} x_j(k) S_k$$
$$= \sum_{k=1}^{r} x_j(k) \sqrt{\phi(k)} = \sum_{k=1}^{r} \theta_j(k) \sqrt{\phi(k)} \sqrt{\phi(k)} = \sum_{k=1}^{r} \theta_j(k) \phi(k).$$

By construction:

$$m_{ij}(p, q) = \sum_{k=1}^{r} x_i(k) x_j(k) = \sum_{k=1}^{r} \sqrt{\phi(k)} \theta_i(k) (\sqrt{\phi(k)} \theta_j(k)) = \sum_{k=1}^{r} \phi(k) \theta_i(k) \theta_j(k).$$

By Lemma 2, this implies that μ is a CIID representation of (p, q). □

Proof of Proposition 5. If q and $\hat{q}$ are related as in equation (15), then $m_{ij}(p, q) = p_j q_i^{(j)} = p_j \left(\alpha \mathbb{I}_{i=j} + (1 - \alpha) \hat{q}_i^{(j)} \right) = \alpha p_j \mathbb{I}_{i=j} + (1 - \alpha) m_{ij}(p, \hat{q})$. Therefore $M(p, q) = \alpha D + (1 - \alpha) M(p, \hat{q})$ where D is the diagonal matrix with $d_{ii} = p_i$. By Example 2.1 in Berman and Shaked-Monderer [2003] D is completely positive. By our Theorem 2, $M(p, \hat{q})$ is also completely positive. By Theorem 2.2 of Berman and Shaked-Monderer [2003], $M(p, q)$ is also completely positive. Therefore (p, q) has a CIID representation by our Theorem 2. □

Proof of Proposition 6. Fix any $\alpha \in (0, 1)$ and let $\hat{q}^{(i)} = \alpha e_i + (1 - \alpha) q^{(i)}$. By Theorem 2, $M(p, q)$ is completely positive. Observe that $M(p, \hat{q}) = \alpha D + (1 - \alpha) M(p, q)$, where D is the diagonal matrix with $d_{ii} = p_i$. Since D is completely positive by Example 2.1 of Berman and Shaked-Monderer [2003], $M(p, \hat{q})$ is a convex combination of completely positive matrices, so by Theorem 2.2 of Berman and Shaked-Monderer [2003] it is also completely positive. Therefore, $(p, \hat{q})$ has a CIID representation by our Theorem 2, so (p, q) has a proportional-reversal representation. □

Proof of Lemma 3. Suppose (p, q) has a CIID representation with measure μ. We will show that the two stated properties hold.

Pairwise exchangeability. Let $h_t \in Y^t$ satisfy $P_{\mathbf{q}}(h_t) > 0$, where $t \leqslant T - 2$, and let $i, j \in Y$. Let $\mu(\cdot \mid h_t)$ denote the posterior distribution over θ after history h_t.

If $q_i^{(h_t)} > 0$, Bayes' rule gives $q_i^{(h_t)} q_j^{(h_t, i)} = \Pr(Y_{t+1} = i, Y_{t+2} = j \mid h_t) = \int_{\Delta(Y)} \theta_i \theta_j \, \mathrm{d}\mu(\theta \mid h_t)$. If $q_i^{(h_t)} = 0$, then $\mathbb{E}_\mu[\theta_i \mid h_t] = 0$. Since $\theta_i \geqslant 0$, this im-

plies $\theta_i = 0$ $\mu(\cdot \mid h_t)$-almost surely, so $q_i^{(h_t)} q_j^{(h_t,i)} = 0 = \int_{\Delta(Y)} \theta_i \theta_j \, \mathrm{d}\mu(\theta \mid h_t)$. Thus, in all cases, $q_i^{(h_t)} q_j^{(h_t,i)} = \int_{\Delta(Y)} \theta_i \theta_j \, \mathrm{d}\mu(\theta \mid h_t)$. Interchanging i and j gives $q_j^{(h_t)} q_i^{(h_t,j)} = \int_{\Delta(Y)} \theta_j \theta_i \, \mathrm{d}\mu(\theta \mid h_t)$, so $q_i^{(h_t)} q_j^{(h_t,i)} = q_j^{(h_t)} q_i^{(h_t,j)}$.

Reinforcement. Let $h_t \in Y^t$ satisfy $P_{\mathbf{q}}(h_t) > 0$, and let $i \in Y$. If $q_i^{(h_t)} = 0$, then the successor history (h_t, i) has probability zero, so the reinforcement condition is not required under Definition 12. Suppose therefore that $q_i^{(h_t)} > 0$. Then (h_t, i) has positive probability, and Bayes' rule gives $q_i^{(h_t,i)} = \dfrac{\mathbb{E}_\mu[\theta_i^2 \mid h_t]}{\mathbb{E}_\mu[\theta_i \mid h_t]}$. Since $q_i^{(h_t)} = \mathbb{E}_\mu[\theta_i \mid h_t]$, we have $q_i^{(h_t,i)} - q_i^{(h_t)} = \dfrac{\mathbb{E}_\mu[\theta_i^2 \mid h_t] - (\mathbb{E}_\mu[\theta_i \mid h_t])^2}{\mathbb{E}_\mu[\theta_i \mid h_t]} = \dfrac{\mathrm{Var}_\mu(\theta_i \mid h_t)}{\mathbb{E}_\mu[\theta_i \mid h_t]} \geqslant 0$. Therefore reinforcement holds. □

Proof of Lemma 4. Consider $\nu \in \mathcal{N}(\mathbf{q})$ such that $\nu + e_j \in \mathcal{N}(\mathbf{q})$ for all $j \in Y$. By pairwise exchangeability, $q_i^{(\nu)} q_j^{(\nu+e_i)} = q_j^{(\nu)} q_i^{(\nu+e_j)}$ for every $i, j \in Y$, so for any $i \in Y$, $q_i^{(\nu)} = \sum_{j \in Y} q_i^{(\nu)} q_j^{(\nu+e_i)} = \sum_{j \in Y} q_j^{(\nu)} q_i^{(\nu+e_j)}$, which is the martingale property. □

Theorem A (Schmüdgen, 2017 Theorems 10.1 and 10.2)**.** Let $n = 2$ and let $(l_0, \ldots, l_T) \in [0,1]^{T+1}$ with $l_0 = 1$. There is a probability measure μ on $[0,1]$ such that $\mathbb{E}_\mu[\theta^k] = l_k$ for all $k \in \{0, \ldots, T\}$ if and only if the associated Hankel matrices $\underline{H}$ and $\overline{H}$ are positive semidefinite.

Proof of Lemma 5. (1) Let $h, h' \in Y^t$ have the same count vector, where $t \leqslant T$. It suffices to show that the probabilities of two histories that differ by one adjacent transposition are equal, since any permutation transforming one ordering into the other can be written as a finite composition of adjacent transpositions.

Write the two histories as $h = (g, i, j, s)$ and $h' = (g, j, i, s)$, where g is the common prefix and s is the common suffix.

If $P_{\mathbf{q}}(g) = 0$, then both $P_{\mathbf{q}}(h)$ and $P_{\mathbf{q}}(h')$ are zero. Suppose instead that $P_{\mathbf{q}}(g) > 0$. By pairwise-exchangeability, $q_i^{(g)} q_j^{(g,i)} = q_j^{(g)} q_i^{(g,j)}$, so $P_{\mathbf{q}}(g, i, j) = P_{\mathbf{q}}(g, j, i)$. The two histories (g, i, j) and (g, j, i) have the same count vector. Thus, by count sufficiency, their forecasts for the remaining suffix s coincide whenever their common probability is positive. If that probability is zero, all extensions have probability zero. Hence $P_{\mathbf{q}}(h) = P_{\mathbf{q}}(h')$ and the moments m_ν are well-defined.

(2) Suppose there exists a probability measure $\mu \in \Delta(\Delta(Y))$ such that for every multi-index ν with $||\nu||_1 \leqslant T$, $m_\nu = \int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta)$. Fix a count vector ν with

$||\nu||_1 < T$ and suppose $m_\nu > 0$. By the definition of implied moments, $q_i^{(\nu)} = \dfrac{m_{\nu+e_i}}{m_\nu}$. Since μ generates the moments $m_{\nu+e_i} = \int_{\Delta(Y)} \theta_i \theta^\nu \, \mathrm{d}\mu(\theta)$ and $m_\nu = \int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta)$. Hence

$$q_i^{(\nu)} = \frac{\int_{\Delta(Y)} \theta_i \theta^\nu \, \mathrm{d}\mu(\theta)}{\int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta)} = \mathbb{E}_\mu[\theta_i \mid \nu] = \int_{\Delta(Y)} \theta_i \, \mathrm{d}\mu(\theta \mid \nu),$$

which is precisely the one-step-ahead forecast generated by the CIID model with prior μ. If $m_\nu = 0$, $\nu \notin \mathcal{N}(\mathbf{q})$, so the conclusion still holds by point (iii) of the definition of a CIID representation.This proves part (2). □

The next condition is motivated by the fact that if the moments $\{m_{r,0}\}_{r=0}^T$ arise from some measure of θ, then for every pair a, b with $a + b \leqslant T$ there is a uniquely determined mixed moment $m_{(a,b)} := \mathbb{E}\big[\theta^a (1-\theta)^b\big]$, since $\theta^a(1-\theta)^b$ is a polynomial of degree $a + b$ in θ.

Definition 17. Forecast $\mathbf{q}$ satisfies *polynomial consistency* if $m_{a,b} = \sum_{i=0}^b \binom{b}{i} (-1)^i m_{i+a,0}$, for all $a, b \geqslant 0$, $a + b \leqslant T$.

Lemma A.2. *If $Y = \{1, 2\}$ and $\mathbf{q}$ satisfies count sufficiency and pairwise exchangeability, then it satisfies polynomial consistency for every $a, b \geqslant 0$ with $a + b \leqslant T$.*

Proof of Theorem 5. Suppose first that $\mathbf{q}$ has a CIID representation with prior μ. Pairwise exchangeability and reinforcement follow from Lemma 3. The implied pure moments satisfy $m_{r,0} = \int_0^1 \theta^r \, \mathrm{d}\mu(\theta), \qquad r = 0, \ldots, T$ so the Hankel matrices are positive semidefinite by Theorem A.

Conversely, suppose that $\mathbf{q}$ satisfies the stated conditions. By Lemma 5, its implied moments are well-defined. Since $m_{0,0} = 1$ and the Hankel matrices are positive semidefinite, Theorem A provides a probability measure μ on $[0, 1]$ such that $\int_0^1 \theta^r \, \mathrm{d}\mu(\theta) = m_{r,0}, r \in \{0, \ldots, T\}$. For every $a, b \geqslant 0$ with $a + b \leqslant T$, the binomial theorem and Lemma A.2 imply

$$\int_0^1 \theta^a (1-\theta)^b \, \mathrm{d}\mu(\theta) = \sum_{i=0}^b (-1)^i \binom{b}{i} \int_0^1 \theta^{a+i} \, \mathrm{d}\mu(\theta) = \sum_{i=0}^b (-1)^i \binom{b}{i} m_{a+i,0} = m_{a,b}.$$

Thus μ matches every implied moment, including those associated with zero-probability histories. Identifying $\theta \in [0, 1]$ with $(\theta, 1-\theta) \in \Delta(Y)$, Lemma 5(2) implies that μ is a CIID representation of $\mathbf{q}$. Forecasts are recovered by moment ratios at histories

with positive probability; at zero-probability histories they are not restricted by the definition of a CIID representation. □

Proof of Proposition 8. Let π denote the agent's prior over the fraction θ of outcome-1 balls in the urn. Rabin's assumption that some urn composition with at least two balls of each outcome has positive prior probability implies $0 < q_1^{(0,0)} < 1$ and makes all conditioning histories below have positive probability.

The first two draws are exchangeable conditional on θ, and hence also unconditionally. Consequently, the two-period forecasts satisfy symmetry: $q_1^{(0,0)}(1-q_1^{(1,0)}) = (1-q_1^{(0,0)})q_1^{(0,1)}$. It follows that $q_2^{(0,1)} - (1-q_1^{(0,0)}) = \frac{q_1^{(0,0)}}{1-q_1^{(0,0)}}(q_1^{(1,0)} - q_1^{(0,0)})$. Thus reinforcement for both outcomes is equivalent to $q_1^{(1,0)} \geqslant q_1^{(0,0)}$. By Theorem 1, the forecasts through period 2 admit a CIID representation if and only if $q_1^{(1,0)} \geqslant q_1^{(0,0)}$.

For the three-period claim, let $P(ijk)$ denote the agent's subjective probability of the ordered sequence (i,j,k). The first two draws are without replacement, and the urn is renewed before the third draw. Therefore, $P(112) = \mathbb{E}_\pi\left[\theta \frac{N\theta-1}{N-1}(1-\theta)\right], P(121) = \mathbb{E}_\pi\left[\theta \frac{N(1-\theta)}{N-1}\theta\right]$. Subtracting gives $P(121) - P(112) = \frac{\mathbb{E}_\pi[\theta(1-\theta)]}{N-1} > 0$. The inequality is strict because the prior assigns positive probability to an urn containing both outcomes.

By the chain rule, $P(112) = q_1^{(0,0)}\, q_1^{(1,0)} q_2^{(2,0)}, P(121) = q_1^{(0,0)}\, q_2^{(1,0)} q_1^{(1,1)}$. Dividing by $q_1^{(0,0)} > 0$, we obtain $q_2^{(1,0)}q_1^{(1,1)} - q_1^{(1,0)}q_2^{(2,0)} = \frac{\mathbb{E}_\pi[\theta(1-\theta)]}{(N-1)q_1^{(0,0)}} > 0$. This violates pairwise exchangeability and by Theorem 5 violates CIID representation. ▫ □

Proof of Theorem 6. First assume $\mathbf{q}$ has a CIID representation. Thus there exists a random $\theta \in \Delta(Y)$ with law μ such that, conditional on θ, the process $(Y_t)_{t \leqslant T}$ is i.i.d. with distribution θ, and the one–step–ahead forecasts induced by this process coincide with $\mathbf{q}$.

Let $m_\nu^* := \int_{\Delta(Y)} \prod_{i \in Y} \theta_i^{\nu_i}\, \mathrm{d}\mu(\theta)$, $||\nu||_1 \leqslant T$. Under the CIID model, Bayes' rule implies that after any history with count vector ν the forecast of $Y_{||\nu||_1+1} = i$ is $q_i(\nu) = \mathbb{E}_\mu[\theta_i \mid \nu] = \frac{\int_{\Delta(Y)} \theta_i \prod_{j \in Y} \theta_j^{\nu_j}\, \mathrm{d}\mu(\theta)}{\int_{\Delta(Y)} \prod_{j \in Y} \theta_j^{\nu_j}, \mathrm{d}\mu(\theta)} = \frac{m^*_{\nu+e_i}}{m^*_\nu}$, whenever $m_\nu^* > 0$. With this, it is immediate that $m_\nu(\mathbf{q}) = m_\nu^* = \int_{\Delta(Y)} \theta^\nu\, \mathrm{d}\mu(\theta)$ for every $\nu \in \mathcal{N}(\mathbf{q})$.

By Lemma B.2, the existence of such a measure μ with $m_\nu = \int \theta^\nu \mathrm{d}\mu$ for all $||\nu||_1 \leqslant T$ implies that, for each $k \leqslant T$, the tensor $M^{(k)}$ is simplex completely positive: there exist an integer $r \geqslant 1$, weights $\gamma_1, \ldots, \gamma_r > 0$ with $\sum_s \gamma_s = 1$, and

points $\pi^{(1)}, \ldots, \pi^{(r)} \in \Delta(Y)$ such that $M^{(k)}_{i_1 \cdots i_k} = m_{\nu(i_1,\ldots,i_k)} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_k}$. [13] Pairwise exchangeability and reinforcement follow from Lemma 3.

Now assume that the count moment tensor $M^{(T)}$ is completely positive. By condition 1 (pairwise exchangeability) and Lemma 5(1), the implied moments $m_\nu(\mathbf{q})$ are well-defined, i.e., m_ν does not depend on the order in which outcomes are accumulated to reach ν. Fix ν with $\|\nu\|_1 < T$ and choose any history $h_\nu \in Y^{\|\nu\|_1}$ with count vector ν. Then $\sum_{i \in Y} m_{\nu+e_i} = \sum_{i \in Y} P_{\mathbf{q}}(h_\nu, i) = P_{\mathbf{q}}(h_\nu) \sum_{i \in Y} q_i^{(h_\nu)} = P_{\mathbf{q}}(h_\nu) = m_\nu$. If $m_\nu = 0$, then $P_{\mathbf{q}}(h_\nu, i) = 0$ for every $i \in Y$, so the same identity holds at null histories. If $m_\nu > 0$, the identity follows from $q^{(h_\nu)} \in \Delta(Y)$. Hence the simplex consistency identities hold for every ν with $\|\nu\|_1 < T$. These are the linear consistency identities (30). By Lemma B.4 $M^{(T)}$ is simplex completely positive, and since the linear consistency conditions hold Lemma B.1 applies: the linear consistency identities together with simplex complete positivity of $M^{(T)}$ imply that every lower-order tensor $M^{(k)}$, $k = 1, \ldots, T-1$, is also simplex completely positive, with the same weights and atoms. Hence all tensors $\{M^{(k)}\}_{k=1}^{T}$ are simplex completely positive, so the hypotheses of Lemma B.2 are satisfied. That lemma then implies there exists a probability measure μ on $\Delta(Y)$ such that $m_\nu = \int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta)$ for all $||\nu||_1 \leqslant T$.

Define a stochastic process by first drawing $\theta \sim \mu$ and then, conditional on θ, drawing $Y_1, \ldots, Y_T$ i.i.d. with distribution θ. Let $(\tilde{p}, \tilde{q})$ denote the associated one–step–ahead forecasts. By construction, $\tilde{p}_i = \Pr(Y_1 = i) = \int_{\Delta(Y)} \theta_i \, \mathrm{d}\mu(\theta) = m_{e_i}$. After observing a history with count vector ν (with $||\nu||_1 < T$), Bayes' rule implies that the posterior on θ has density proportional to $\theta^\nu \mathrm{d}\mu(\theta)$. Hence the next-step forecast satisfies $\tilde{q}_i(\nu) = \mathbb{E}_\mu[\theta_i \mid \nu] = \frac{\int_{\Delta(Y)} \theta_i \theta^\nu \, \mathrm{d}\mu(\theta)}{\int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta)} = \frac{m_{\nu+e_i}}{m_\nu}$ from $m_\nu = \int \theta^\nu \mathrm{d}\mu(\theta)$. Since the forecast-defined moments satisfy $q_i(\nu) = \frac{m_{\nu+e_i}}{m_\nu}$, the CIID process generated by μ reproduces the original forecast system at all histories with $m_\nu > 0$. When $m_\nu = 0$, the value of $q(\nu)$ does not affect the induced measure. Hence the forecast system admits a CIID representation. □

[13] The map $\theta \mapsto \theta^{\otimes k}$ is continuous on the compact set $\Delta(Y)$, so its image is compact, and its convex hull is compact. Carathéodory's theorem therefore guarantees that each $M^{(k)}$ admits a finite simplex completely positive decomposition.

B For Online Publication

B.1 Diagonal Dominance

Example 5. *Suppose* $n = 5$ *and that the forecasts are* $p = (3/109, 53/218, 53/218, 53/218, 53/218)$, $q^{(1)} = (\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$, *and* $q_1^{(i)} = 1/53$, $q_i^{(i)} = 101/106$ *and* $q_j^{(i)} = 1/106$ *for all* $i, j \in Y \backslash \{1\}$ *with* $i \neq j$. *Then*

$$M(p,q) = \frac{1}{436}\begin{pmatrix} 4 & 2 & 2 & 2 & 2 \\ 2 & 101 & 1 & 1 & 1 \\ 2 & 1 & 101 & 1 & 1 \\ 2 & 1 & 1 & 101 & 1 \\ 2 & 1 & 1 & 1 & 101 \end{pmatrix}$$

which is not diagonal dominant. The associated row-ratio matrix is

$$R(p,q) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{101} & 0 & \frac{1}{101} & \frac{1}{101} & \frac{1}{101} \\ \frac{2}{101} & \frac{1}{101} & 0 & \frac{1}{101} & \frac{1}{101} \\ \frac{2}{101} & \frac{1}{101} & \frac{1}{101} & 0 & \frac{1}{101} \\ \frac{2}{101} & \frac{1}{101} & \frac{1}{101} & \frac{1}{101} & 0 \end{pmatrix}$$

Since the eigenvalues of $R(p,q)$ *are* $\lambda_1 \approx 0.2144$, $\lambda_2 \approx -0.1847$ *and* $\lambda = -\frac{1}{101} \approx -0.0099$ *(with multiplicity 3), the spectral radius of* $R(p,q)$ *is*

$$\rho(R(p,q)) = \approx 0.2144 < 1.$$

B.2 Proof of Corollary 2

Proof of Corollary 2. For $p_j > 0$, let $q^{B(j)} = \int \theta \, \mathrm{d}\mu_j(\theta)$ denote the forecast under exact Bayesian updating. For $p_j = 0$, set $q^{B(j)} = q^{(j)}$. Then (p, q^B) has the CIID representation μ, and its matrix is

$$A = \int_{\Delta(Y)} \theta\theta^\top \, \mathrm{d}\mu(\theta) \in \mathcal{CP}_n.$$

For any subset $S \subseteq Y$, the function $f_S(\theta) = \sum_{i \in S} \theta_i$ takes values in $[0,1]$. Therefore

$$\left| \sum_{i \in S} (q_i^{(j)} - q_i^{B(j)}) \right| = \left| \int f_S \, \mathrm{d}(\widetilde{\mu}_j - \mu(\cdot|j)) \right| \leqslant \eta_j.$$

Here we use the elementary identity $\sup_{0 \leqslant f \leqslant 1} |\int f \, \mathrm{d}(\nu - \rho)| = d_{\mathrm{TV}}(\nu, \rho)$ for probability measures, which follows by writing $f(\theta) = \int_0^1 \mathbf{1}_{\{f(\theta) > t\}} \, \mathrm{d}t$. Taking the supremum over S and using the finite-space formula for total variation yields

$$\|q^{(j)} - q^{B(j)}\|_1 \leqslant 2\eta_j.$$

Substitute this inequality into (11) to obtain (12). Substituting this inequality into (11) gives the first inequality in (12). The second follows from

$$\left(\sum_{j \in Y} p_j^2 \eta_j^2 \right)^{1/2} \leqslant \sum_{j \in Y} p_j \eta_j,$$

since the Euclidean norm of a nonnegative vector is at most the sum of its coordinates. □

B.3 Grether updating

Proof of Proposition 7. CIID rationalizability implies symmetry by the main theorem. For the converse, let $\delta := a - b$ and $z_i := p_i(w_i - \lambda)$. Since $p = b + \lambda\delta$, we have $q_j^{(i)} = p_j + (w_i - \lambda)\delta_j$, and hence $M^G = pp^\top + \delta z^\top$.

If $a = b$, then $q^{(i)} = p$ for every i, and the prior concentrated on p is a CIID representation. Suppose, therefore, that $a \neq b$. Symmetry is equivalent to $\delta z^\top = z\delta^\top$. Choosing h with $\delta_h \neq 0$ and comparing the (i,h) entries shows that $z = \kappa\delta$ for some scalar κ; the converse is immediate. We next show that $\kappa > 0$. Because $\sum_i \delta_i = 0$ and $\delta \neq 0$, there are outcomes h and ℓ such that $\delta_h > 0 > \delta_\ell$. Hence $a_h/b_h > 1 > a_\ell/b_\ell$. Since $c > 0$, Grether's posterior-odds formula implies $w_h > w_\ell$. On the other hand, $z = \kappa\delta$ implies

$$w_i - \lambda = \kappa \frac{\delta_i}{p_i}.$$

If $\kappa < 0$, then $w_h < \lambda < w_\ell$, contradicting $w_h > w_\ell$. If $\kappa = 0$, then $z = 0$ and

$w_h = w_\ell = \lambda$, yielding the same contradiction, so $\kappa > 0$.

It remains to construct a CIID representation. Symmetry gives $M^G = pp^\top + \kappa\delta\delta^\top$. Define $U := \min_{\delta_i<0} p_i/(-\delta_i)$ and $S := \min_{\delta_i>0} p_i/\delta_i$. For every pair with $\delta_i < 0 < \delta_j$, nonnegativity of $M^G_{ij} = p_i p_j + \kappa\delta_i\delta_j$ implies $\kappa \leqslant [p_i/(-\delta_i)][p_j/\delta_j]$; hence $\kappa \leqslant US$.

Set $t := U$, $s := \kappa/U$, $\pi^+ := p + t\delta$, and $\pi^- := p - s\delta$. Because $0 < s \leqslant S$, the definitions of U and S imply that $\pi^+, \pi^- \in \Delta(Y)$. Assign probabilities $\gamma^+ := s/(s+t)$ and $\gamma^- := t/(s+t)$ to π^+ and π^-. Then $\gamma^+ t = \gamma^- s$, so this prior has mean p, and

$$\gamma^+\pi^+\pi^{+\top} + \gamma^-\pi^-\pi^{-\top} = pp^\top + st\,\delta\delta^\top = pp^\top + \kappa\delta\delta^\top = M^G.$$

Its first moments are therefore p_i and its second moments are $p_i q_j^{(i)}$. Its Bayesian forecast after observing i is consequently $q^{(i)}$, so it is a CIID representation. □

We finally specialize to two outcomes, labeled 0 and 1. Write $a := \Pr(1 \mid A)$, $b := \Pr(1 \mid B)$, and $p := \lambda a + (1-\lambda)b$. Let q_1 denote the forecast of outcome 1 after observing 1, and q_0 the forecast of outcome 1 after observing 0. Symmetry is then the single condition $p(1-q_1) = (1-p)q_0$.

Define $R := \lambda/(1-\lambda)$, $L_1 := (a/b)^c$, $L_0 := ((1-a)/(1-b))^c$, $x := R^d$, $K := pL_1 + (1-p)L_0$, and $B := K - \lambda(L_1 + L_0)$.

Proposition 10. *Suppose $0 < b \leqslant a < 1$, $\lambda \in (0,1)$, and $c, d > 0$. If $a = b$, symmetry holds for every λ, c, and d. If $a > b$, symmetry holds if and only if*

$$(1-\lambda)L_1L_0x^2 + Bx - \lambda = 0, \qquad x = \left(\frac{\lambda}{1-\lambda}\right)^d.$$

Equivalently, symmetry holds if and only if $R^d = x^*$, where

$$x^* = \frac{-B + \sqrt{B^2 + 4\lambda(1-\lambda)L_1L_0}}{2(1-\lambda)L_1L_0}.$$

Proof of Proposition 10. If $a = b$, then $p = q_1 = q_0 = a$, so symmetry holds. Suppose $a > b$. Grether's rule gives $w_1 = xL_1/(1+xL_1)$ and $w_0 = xL_0/(1+xL_0)$. The symmetry condition is equivalent to $p = pq_1 + (1-p)q_0$. Substituting $q_y = b + (a-b)w_y$ and $p = b + (a-b)\lambda$, and canceling $a - b > 0$, shows that symmetry is equivalent to

$$p\frac{xL_1}{1+xL_1} + (1-p)\frac{xL_0}{1+xL_0} = \lambda.$$

Multiplying by $(1 + xL_1)(1 + xL_0)$ and collecting terms yields the stated quadratic. Its leading coefficient is positive and its constant term is negative, so it has exactly one positive root, namely x^*. Because $x = R^d > 0$, symmetry holds exactly when $R^d = x^*$. □

If $\lambda \neq \frac{1}{2}$, the condition can equivalently be written $d = \log(x^*)/\log R$. If $\lambda = \frac{1}{2}$, then $R^d = 1$ for every $d > 0$, so symmetry is independent of d and holds exactly when $x^* = 1$.

The following example shows that Grether updating need not satisfy symmetry, even though it satisfies reinforcement.

Example 6. *Let $\lambda = \frac{4}{5}$, $a = \frac{4}{5}$, $b = \frac{1}{5}$, $c = 1$, and $d = \frac{1}{2}$. The initial forecast is $p = \frac{4}{5}\frac{4}{5} + \frac{1}{5}\frac{1}{5} = \frac{17}{25}$. The prior odds are 4. After observing outcome 1, the posterior odds are $4^{1/2}(a/b) = 8$, so $w_1 = 8/9$. After observing outcome 0, they are $4^{1/2}(1 - a)/(1 - b) = 1/2$, so $w_0 = 1/3$. Hence $q_1 = \frac{1}{5} + \frac{3}{5}\frac{8}{9} = \frac{11}{15}$ and $q_0 = \frac{1}{5} + \frac{3}{5}\frac{1}{3} = \frac{2}{5}$. Symmetry would require $p(1 - q_1) = (1 - p)q_0$, but*

$$p(1 - q_1) = \frac{68}{375} \qquad \text{and} \qquad (1 - p)q_0 = \frac{48}{375}.$$

Thus symmetry fails, and the forecasts are not CIID-rationalizable. They nevertheless satisfy reinforcement: $q_1 = \frac{11}{15} > \frac{17}{25} > \frac{2}{5} = q_0$.

B.4 Lemma A.1

Proof. All expectations below are taken over the independent draws $\theta^{(1)}, \ldots, \theta^{(k)}$ from μ. Let $X_s = \theta^{(s)}\theta^{(s)\top}$, so that $\mathbb{E}[X_s] = M(p, q)$ and $M_k - M(p, q) = k^{-1}\sum_{s=1}^{k}(X_s - M(p, q))$. By Jensen's inequality,

$$\mathbb{E}\|M_k - M(p, q)\|_F \leqslant \left(\mathbb{E}\|M_k - M(p, q)\|_F^2\right)^{1/2}. \tag{25}$$

To compute the expected squared norm, observe that for each fixed pair (i, j), the variables $Z_{s,ij} := (X_s)_{ij} - m_{ij}(p, q)$ are independent across s and have mean zero, so when $s \neq t$, $\mathbb{E}[Z_{s,ij}Z_{t,ij}] = \mathbb{E}[Z_{s,ij}]\,\mathbb{E}[Z_{t,ij}] = 0$. Since the squared Frobenius norm

is the sum of the squared entries,

$$\mathbb{E}\|M_k - M(p,q)\|_F^2 = \frac{1}{k^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{k}\sum_{t=1}^{k}\mathbb{E}[Z_{s,ij}Z_{t,ij}] = \frac{1}{k^2}\sum_{s=1}^{k}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[Z_{s,ij}^2] = \frac{1}{k}\,\mathbb{E}\|X_1 - M(p,q)\|_F^2, \tag{26}$$

where the second equality drops the zero cross terms and the third uses the fact that the X_s are identically distributed.

Next, for every $i, j \in \{1, ..., n\}$, using $\mathbb{E}[(X_1)_{ij}] = m_{ij}(p,q)$ gives

$$\mathbb{E}\big[((X_1)_{ij} - m_{ij}(p,q))^2\big] = \mathbb{E}\big[(X_1)_{ij}^2\big] - 2m_{ij}(p,q)\mathbb{E}[(X_1)_{ij}] + m_{ij}(p,q)^2 = \mathbb{E}\big[(X_1)_{ij}^2\big] - m_{ij}(p,q)^2.$$

Summing over entries therefore yields

$$\mathbb{E}\|X_1 - M(p,q)\|_F^2 = \mathbb{E}\|X_1\|_F^2 - \|M(p,q)\|_F^2 \leqslant \mathbb{E}\|X_1\|_F^2. \tag{27}$$

Moreover,

$$\|X_1\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\big(\theta_i^{(1)}\theta_j^{(1)}\big)^2 = \left(\sum_{i=1}^{n}(\theta_i^{(1)})^2\right)\left(\sum_{j=1}^{n}(\theta_j^{(1)})^2\right) = \|\theta^{(1)}\|_2^4 \leqslant \|\theta^{(1)}\|_1^4 = 1, \tag{28}$$

where the inequality holds because $\theta^{(1)} \in \Delta(Y)$. Combining equations (25), (26), (27), and (28), we obtain $\mathbb{E}\|M_k - M(p,q)\|_F \leqslant k^{-1/2}$, which proves the first bound. In particular, there exist draws $\theta^{(1)}, \ldots, \theta^{(k)}$ for which $\|M_k - M(p,q)\|_F \leqslant k^{-1/2}$. □

B.5 Uniqueness and Multiplicity

Let $\mathrm{Pos}([0,1])_T$ denote the cone of real polynomials of degree at most T that are nonnegative on the interval $[0,1]$, and for any sequence $(m_0, \ldots, m_T) \in \mathbb{R}^{T+1}$ let L_m be the linear functional on $\mathbb{R}[x]_{\leqslant T}$ defined by $L_m(x^j) = m_j$ for all $j \in \{0, ..., T\}$. We will say that sequence $(m_0, \ldots, m_T)$ is *interior* if $L_m(p) > 0$ for every nonzero $p \in \mathrm{Pos}([0,1])_T$. Recall that when $n = 2$, if $\mathbf{q}$ satisfies count sufficiency and pairwise exchangeability, then by Lemma A.2 and the argument in the proof of Theorem 5, a measure on $[0,1]$ matches the implied mixed moments $m_{a,b}$ if and only if it matches the pure moments $m_{r,0}$ for $r \leqslant T$.

Theorem 7. *Let $T = 2j-1$ for an integer $j \geqslant 1$, and suppose that $m = (m_0, \ldots, m_T)$*

is interior and $m_0 = 1$. *There exists a unique probability measure* $\mu \in \Delta([0,1])$ *with exactly* j *support points such that* $\mathbb{E}_\mu[\theta^t] = m_t$ *for every* $t = 0, \ldots, T$. *No representing measure has fewer than* j *support points.*

Theorem 8. *Let* $T = 2k$ *for an integer* $k \geqslant 1$, *and suppose that the associated moment sequence* $m = (m_0, \ldots, m_{2k})$ *is interior, with* $m_0 = 1$. *Then there exists a one-parameter family of distinct rationalizing priors, each supported on exactly* $k+1$ *points, that all represent the same truncated moment sequence* $(m_0, \ldots, m_{2k})$.

The following moment results will be used below:

Theorem A (Schmüdgen, 2017 Theorem 10.7)**.** Let $k \in \mathbb{N}$ and let $m = (m_0, \ldots, m_{2k})$ with $m_0 = 1$. The following are equivalent:

(i) There exists $\mu \in \Delta([0,1])$ such that

$$m_j = \int_0^1 x^j \, \mathrm{d}\mu(x), \qquad \forall j \in \{0, \ldots, 2k\}. \tag{29}$$

(ii) $L_m(p) \geqslant 0$ for every $p \in \mathrm{Pos}([0,1])_{2k}$.

Moreover, m lies in the interior of the truncated moment cone on $[0,1]$ if and only if $L_m(p) > 0$ for every nonzero $p \in \mathrm{Pos}([0,1])_{2k}$.

This theorem implies that when the Hankel matrices of the implied moments are (strictly) positive definite, the implied moment sequence $(m_{(0,0)}, m_{(1,0)}, \ldots, m_{(T,0)})$ is interior.

Proof of Theorem 7. Whenever $\deg f + \deg g \leqslant T$, write $\langle f, g \rangle := L_m(fg)$. On $\mathbb{R}[x]_{\leqslant j-1}$ this defines a positive definite bilinear form: for every nonzero f in this space, f^2 is a nonzero nonnegative polynomial of degree at most $2j-2$, so interiority gives $\langle f, f \rangle = L_m(f^2) > 0$.

Claim 2. *There exists a unique monic polynomial* $P_j(x)$ *of degree* j *that is orthogonal to all polynomials of degree at most* $j-1$.

Proof. We first construct, by induction, monic polynomials P_l of degree l, for $l = 0, \ldots, j$, each orthogonal to every polynomial of degree at most $l-1$.

For $l = 0$, set $P_0(x) = 1$. Suppose that $P_0, \ldots, P_{l-1}$ have been constructed. They form an orthogonal basis for $\mathbb{R}[x]_{\leqslant l-1}$. Set

$$P_l(x) = x^l - \sum_{i=0}^{l-1} \frac{\langle x^l, P_i \rangle}{\langle P_i, P_i \rangle} P_i(x).$$

All products appearing here have degree at most $2j-1$, and every denominator is strictly positive by interiority. The polynomial P_l is monic of degree l, and orthogonality of $P_0, \ldots, P_{l-1}$ gives $\langle P_l, P_i \rangle = 0$ for every $i < l$. Thus P_l is orthogonal to every polynomial of degree at most $l-1$. We next establish uniqueness. Suppose there are two monic polynomials of degree j, $P_j(x)$ and $\tilde{P}_j(x)$, both orthogonal to $\mathbb{R}[x]_{\leqslant j-1}$. Let $D(x) = P_j(x) - \tilde{P}_j(x)$. Since both are monic of degree j, the x^j terms cancel, meaning the degree of $D(x)$ is at most $j-1$. Thus, $D \in \mathbb{R}[x]_{\leqslant j-1}$.

Because P_j and $\tilde{P}_j$ are orthogonal to all polynomials in $\mathbb{R}[x]_{\leqslant j-1}$, they are both orthogonal to $D(x)$: $\langle D, D \rangle = \langle P_j - \tilde{P}_j, D \rangle = \langle P_j, D \rangle - \langle \tilde{P}_j, D \rangle = 0 - 0 = 0$. Since $\deg D \leqslant j-1$, the polynomial D^2 has degree at most $2j-2$ and is nonnegative on $[0,1]$. Suppose that $D \not\equiv 0$. Then D^2 is a nonzero element of $\mathrm{Pos}([0,1])_{2j-1}$, so interiority implies $\langle D, D \rangle = L_m(D^2) > 0$. This contradicts $\langle D, D \rangle = 0$. Hence $D \equiv 0$, and therefore $P_j = \tilde{P}_j$, proving uniqueness. □

We next show that P_j has j distinct roots in $(0,1)$. Let r be the number of distinct roots in $(0,1)$ at which P_j changes sign. If $r < j$, let R be the product of the corresponding linear factors, choosing its overall sign so that $P_j R \geqslant 0$ on $[0,1]$. Then $P_j R$ is nonzero and has degree $j + r \leqslant 2j-1$, so interiority gives $L_m(P_j R) > 0$. But $\deg R < j$, so orthogonality gives $L_m(P_j R) = 0$, a contradiction. Hence $r = j$, and all roots of P_j are distinct and lie in $(0,1)$. Denote them by $p_1, \ldots, p_j$. For $k = 1, \ldots, j$, define

$$\ell_k(x) := \prod_{i \neq k} \frac{x - p_i}{p_k - p_i}, \qquad \lambda_k := L_m(\ell_k).$$

For any polynomial f of degree at most $2j-1$, polynomial division gives $f = P_j S + R$, where $\deg S, \deg R \leqslant j-1$. Orthogonality and Lagrange interpolation therefore imply

$$L_m(f) = L_m(R) = \sum_{k=1}^{j} \lambda_k R(p_k) = \sum_{k=1}^{j} \lambda_k f(p_k).$$

Taking $f = \ell_k^2$ gives $\lambda_k = L_m(\ell_k^2) > 0$ by interiority. Taking $f = 1$ gives $\sum_{k=1}^{j} \lambda_k = m_0 = 1$. Thus $\sum_{k=1}^{j} \lambda_k \delta_{p_k}$ is a probability measure with exactly j support points that matches $m_0, \ldots, m_{2j-1}$.

If a representing measure had $r < j$ support points $t_1, \ldots, t_r$, let $A(x) = \prod_{i=1}^{r}(x - t_i)$. Then $L_m(A^2) = 0$, contradicting interiority because A^2 is nonzero, nonnegative, and has degree $2r \leqslant 2j - 2$.

To establish uniqueness, suppose that another j-point representing measure $\widetilde{\mu}$ has support $\{q_1, \ldots, q_j\}$ and weights $w_1, \ldots, w_j$. Let $Q(x) = \prod_{k=1}^{j}(x - q_k)$. For every polynomial R of degree at most $j - 1$, moment matching gives

$$\langle Q, R \rangle = L_m(QR) = \int Q(x)R(x)\,\mathrm{d}\widetilde{\mu}(x) = 0.$$

By the claim, $Q = P_j$, so after relabeling the support points we have $q_k = p_k$ for every k. Consequently,

$$w_k = \int \ell_k(x)\,\mathrm{d}\widetilde{\mu}(x) = L_m(\ell_k) = \lambda_k.$$

Thus the j-point representing measure is unique. □

Proof of Theorem 8. Let $m = (m_0, \ldots, m_{2k})$ be an interior truncated moment sequence on $[0, 1]$, with $m_0 = 1$, and let L_m denote its associated linear functional.

Step 1 Suppose toward a contradiction that there is a rationalizing prior μ for m supported on $\{t_1, \ldots, t_r\} \subset [0, 1]$ with $r \leqslant k$. Define the polynomial $l(x) := \prod_{j=1}^{r}(x - t_j)$, which has degree $r \leqslant k$ and vanishes at every support point of μ. Then l^2 is a nonzero polynomial of degree at most $2k$ that is nonnegative on $[0, 1]$, and $L_m(l^2) = \int_0^1 l(x)^2\,\mathrm{d}\mu(x) = 0$, contradicting interiority. Hence no rationalizing prior with k or fewer support points exists.

Step 2. We first show that m admits an interior extension to order $2k + 1$. Let U denote the uniform probability measure on $[0, 1]$, and let

$$u := \left(1, \frac{1}{2}, \ldots, \frac{1}{2k+1}\right)$$

be its moment vector through order $2k$. By Theorem A, m lies in the interior of the

moment cone. Hence, for sufficiently small $\varepsilon \in (0,1)$, the vector

$$\widetilde{m} := \frac{m - \varepsilon u}{1 - \varepsilon}$$

still belongs to that cone and has zeroth moment equal to one. Choose a probability measure μ representing $\widetilde{m}$, and define

$$\rho := (1 - \varepsilon)\mu + \varepsilon U.$$

Then ρ represents m and has full support on $[0,1]$.

For $s \in \mathbb{R}$, write

$$m(s) := (m_0, \ldots, m_{2k}, s), \qquad s_0 := \int_0^1 x^{2k+1} \, \mathrm{d}\rho(x).$$

For every nonzero polynomial P of degree at most $2k+1$ that is nonnegative on $[0,1]$, we have

$$L_{m(s_0)}(P) = \int_0^1 P(x) \, \mathrm{d}\rho(x) \geqslant \varepsilon \int_0^1 P(x) \, \mathrm{d}x > 0.$$

Thus $m(s_0)$ is interior.

This strict positivity persists for s in an open interval around s_0. Indeed, consider the compact set

$$\mathcal{K} := \left\{ P(x) = \sum_{j=0}^{2k+1} a_j x^j : P \geqslant 0 \text{ on } [0,1], \ \sum_{j=0}^{2k+1} a_j^2 = 1 \right\}.$$

Every polynomial in $\mathcal{K}$ is nonzero, so continuity and compactness give

$$c := \min_{P \in \mathcal{K}} L_{m(s_0)}(P) > 0.$$

Since $|a_{2k+1}| \leqslant 1$ for $P \in \mathcal{K}$,

$$L_{m(s)}(P) = L_{m(s_0)}(P) + (s - s_0) a_{2k+1} \geqslant c - |s - s_0| > 0$$

whenever $|s - s_0| < c$. By positive rescaling, the same strict positivity holds for every nonzero nonnegative polynomial of degree at most $2k+1$. Consequently, $m(s)$

is interior for every

$$s \in I := (s_0 - c, s_0 + c).$$

For each $s \in I$, apply Theorem 7 with $j = k + 1$. Since $2j - 1 = 2k + 1$, that theorem provides a unique probability measure μ_s with exactly $k + 1$ support points such that

$$\int_0^1 x^t \, \mathrm{d}\mu_s(x) = m_t \quad \text{for } t = 0, \ldots, 2k, \qquad \int_0^1 x^{2k+1} \, \mathrm{d}\mu_s(x) = s.$$

Thus every μ_s represents the original truncated moment sequence m. Moreover, if $s \neq s'$, then $\mu_s \neq \mu_{s'}$, because their moments of order $2k + 1$ differ. Therefore $\{\mu_s : s \in I\}$ is a one-parameter family of distinct rationalizing priors, each supported on exactly $k + 1$ points. □

B.6 Ancillary Results for Multiple Periods and Outcomes

Proof of Lemma A.2. By Lemma 5, the implied moments are well-defined. Fix $a, b \geqslant 0$ with $a + b < T$, and choose a history h containing a realizations of outcome 1 and b realizations of outcome 2. Then

$$\begin{aligned} m_{a+1,b} + m_{a,b+1} &= P_{\mathbf{q}}(h, 1) + P_{\mathbf{q}}(h, 2) \\ &= P_{\mathbf{q}}(h)\big(q_1^{(h)} + q_2^{(h)}\big) \\ &= P_{\mathbf{q}}(h) \\ &= m_{a,b}. \end{aligned}$$

This identity holds also when $P_{\mathbf{q}}(h) = 0$, since both extensions then have probability zero. Consequently,

$$m_{a,b+1} = m_{a,b} - m_{a+1,b}.$$

We now proceed by induction on b. For $b = 0$, polynomial consistency is immedi-

ate. Suppose it holds for b and every a with $a+b \leqslant T$. For any a with $a+b+1 \leqslant T$,

$$\begin{aligned} m_{a,b+1} &= m_{a,b} - m_{a+1,b} \\ &= \sum_{i=0}^{b} (-1)^i \binom{b}{i} m_{a+i,0} - \sum_{i=0}^{b} (-1)^i \binom{b}{i} m_{a+i+1,0} \\ &= \sum_{i=0}^{b+1} (-1)^i \binom{b+1}{i} m_{a+i,0}, \end{aligned}$$

where the last equality follows from Pascal's identity. Thus polynomial consistency holds for every $a, b \geqslant 0$ with $a + b \leqslant T$. □

Proof of Proposition 9. Let θ denote the latent probability of outcome 1. The two-period forecasts imply $\mathbb{E}[\theta] = q_1^{(0,0)}$ and $\mathbb{E}[\theta^2] = q_1^{(0,0)} q_1^{(1,0)}$. Write $t := \mathbb{E}[\theta^3]$. Since $q_1^{(2,0)} = t/(q_1^{(0,0)} q_1^{(1,0)})$, set $x := q_1^{(2,0)}$, so that $t = q_1^{(0,0)} q_1^{(1,0)} x$. By Cauchy–Schwarz, $\mathbb{E}[\theta^2]^2 \leqslant \mathbb{E}[\theta]\mathbb{E}[\theta^3]$, hence $(q_1^{(0,0)} q_1^{(1,0)})^2 \leqslant {q_1^{(0,0)}}^2 q_1^{(1,0)} x$ and therefore $x \geqslant q_1^{(1,0)}$. Applying Cauchy–Schwarz to $\sqrt{1-\theta}$ and $\theta\sqrt{1-\theta}$ gives $\mathbb{E}[\theta(1-\theta)]^2 \leqslant \mathbb{E}[1-\theta]\mathbb{E}[\theta^2(1-\theta)]$. Since $\mathbb{E}[\theta(1-\theta)] = q_1^{(0,0)}(1-q_1^{(1,0)})$ and $\mathbb{E}[\theta^2(1-\theta)] = q_1^{(0,0)} q_1^{(1,0)}(1-x)$, this yields $x \leqslant 1 - \frac{q_1^{(0,0)}(1-q_1^{(1,0)})^2}{(1-q_1^{(0,0)})q_1^{(1,0)}}$. The remaining forecasts are determined by the same third moment: $q_1^{(1,1)} = \mathbb{E}[\theta^2(1-\theta)]/\mathbb{E}[\theta(1-\theta)] = q_1^{(1,0)}(1-x)/(1-q_1^{(1,0)})$ and $q_1^{(0,2)} = \mathbb{E}[\theta(1-\theta)^2]/\mathbb{E}[(1-\theta)^2] = q_1^{(0,0)}(1-2q_1^{(1,0)} + q_1^{(1,0)} x)/(1-2q_1^{(0,0)} + q_1^{(0,0)} q_1^{(1,0)})$. It remains to show sharpness. The lower endpoint is generated by $\mu_L = (1 - q_1^{(0,0)}/q_1^{(1,0)})\delta_0 + (q_1^{(0,0)}/q_1^{(1,0)})\delta_{q_1^{(1,0)}}$, which has first two moments $q_1^{(0,0)}$ and $q_1^{(0,0)} q_1^{(1,0)}$ and satisfies $\mathbb{E}_{\mu_L}[\theta^3] = q_1^{(0,0)}(q_1^{(1,0)})^2$. For the upper endpoint, let $b := q_1^{(0,0)}(1-q_1^{(1,0)})/(1-q_1^{(0,0)})$ and take $\mu_U = ((1-q_1^{(0,0)})/(1-b))\delta_b + ((q_1^{(0,0)}-b)/(1-b))\delta_1$. This measure also has first two moments $q_1^{(0,0)}$ and $q_1^{(0,0)} q_1^{(1,0)}$, and its third moment is $q_1^{(0,0)} q_1^{(1,0)} - (q_1^{(0,0)})^2(1-q_1^{(1,0)})^2/(1-q_1^{(0,0)})$, which yields the upper bound on x. Every convex combination of μ_L and μ_U preserves the first two moments and generates an intermediate third moment, so every x in the stated interval is attainable. □

Lemma B.1. *Let $T \geqslant 2$, $Y = \{1, \ldots, n\}$, and let $\mathbf{q}$ be a forecast satisfying count sufficiency. Suppose that the implied moments satisfy the simplex identities*

$$\sum_{i \in Y} m_{\nu + e_i} = m_\nu \qquad \text{for all } ||\nu||_1 \leqslant T-1, \tag{30}$$

and that the order-T count-moment tensor $M^{(T)}$ is simplex completely positive: there

exist $r \in \mathbb{N}$, *weights* $\gamma_s > 0$ *with* $\sum_{s=1}^{r} \gamma_s = 1$, *and vectors* $\pi^{(s)} \in \Delta(Y)$ *such that*

$$M^{(T)}_{i_1 \cdots i_T} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_T} \qquad \forall (i_1, \ldots, i_T) \in Y^T, \tag{31}$$

where for each $(i_1, \ldots, i_T)$ *the entry* $M^{(T)}_{i_1 \cdots i_T}$ *equals* $m_{\nu(i_1,\ldots,i_T)}$ *with* $\nu(i_1, \ldots, i_T)$ *the corresponding count vector. Then, for every* $k \in \{1, \ldots, T-1\}$, *the order-*k *tensor* $M^{(k)}$ *is also simplex completely positive with the same factors and weights:*

$$M^{(k)}_{i_1 \cdots i_k} = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_k} \qquad \forall (i_1, \ldots, i_k) \in Y^k.$$

Proof. The simplex complete positivity assumption (31) at order T means exactly that there is a discrete probability measure μ on $\Delta(Y)$ of the form

$$\mu := \sum_{s=1}^{r} \gamma_s \, \delta_{\pi^{(s)}},$$

such that, for all $(i_1, \ldots, i_T) \in Y^T$,

$$M^{(T)}_{i_1 \cdots i_T} = \int_{\Delta(Y)} \theta_{i_1} \cdots \theta_{i_T} \, \mathrm{d}\mu(\theta).$$

Equivalently, for every count vector ν with $||\nu||_1 = T$,

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta), \tag{32}$$

where $\theta^\nu := \prod_{i \in Y} \theta_i^{\nu_i}$.

We claim that the same representation holds for all lower orders:

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta) \qquad \text{for all } \nu \in \mathbb{N}^n \text{ with } ||\nu||_1 \leqslant T. \tag{33}$$

We prove this by backward induction on $||\nu||_1$.

For $||\nu||_1 = T$ this is exactly (32). Suppose now that (33) holds for all η with $||\eta||_1 = k + 1$, where $1 \leqslant k \leqslant T - 1$, and fix any ν with $||\nu||_1 = k$. By the simplex

identity (30),

$$m_\nu = \sum_{i \in Y} m_{\nu + e_i}.$$

By the induction hypothesis, for each $i \in Y$ we have

$$m_{\nu+e_i} = \int_{\Delta(Y)} \theta^{\nu+e_i} \, \mathrm{d}\mu(\theta) = \int_{\Delta(Y)} \theta_i \, \theta^\nu \, \mathrm{d}\mu(\theta).$$

Summing over i yields

$$m_\nu = \sum_{i \in Y} m_{\nu+e_i} = \int_{\Delta(Y)} \Big(\sum_{i \in Y} \theta_i \Big) \theta^\nu \, \mathrm{d}\mu(\theta).$$

Because each $\pi^{(s)} \in \Delta(Y)$, every θ in the support of μ satisfies $\sum_{i \in Y} \theta_i = 1$. Hence

$$\sum_{i \in Y} \theta_i = 1 \quad \mu\text{-a.s.},$$

and therefore

$$m_\nu = \int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta),$$

which is (33) for $||\nu||_1 = k$. This completes the backward induction and proves (33) for all $||\nu||_1 \leqslant T$.

Finally, fix any $k \in \{1, \dots, T-1\}$ and any $(i_1, \dots, i_k) \in Y^k$ with induced count vector $\nu(i_1, \dots, i_k)$. Then

$$M^{(k)}_{i_1 \cdots i_k} = m_{\nu(i_1,\dots,i_k)} = \int_{\Delta(Y)} \theta_{i_1} \cdots \theta_{i_k} \, \mathrm{d}\mu(\theta) = \sum_{s=1}^{r} \gamma_s \, \pi^{(s)}_{i_1} \cdots \pi^{(s)}_{i_k},$$

which shows that $M^{(k)}$ is simplex completely positive, with the same factors $\{\pi^{(s)}\}$ and weights $\{\gamma_s\}$ as $M^{(T)}$. □

Lemma B.2. *Let $\{m_\nu\}_{||\nu||_1 \leqslant T}$ be a collection of nonnegative numbers with $m_{\mathbf{0}} = 1$. The following are equivalent:*

1. *There exists a probability measure μ on the simplex $\Delta(Y)$ such that*

$$m_\nu = \mathbb{E}_\mu[\prod_{i \in Y} \theta_i^{\nu_i}] = \int_{\Delta(Y)} \prod_{i \in Y} \theta_i^{\nu_i} \, d\mu(\theta) \qquad \forall ||\nu||_1 \leqslant T.$$

2. *The symmetric order-k tensor* $M^{(k)}$ *defined by*

$$M^{(k)}_{i_1\cdots i_k} \;:=\; m_{e_{i_1}+\cdots+e_{i_k}}$$

is simplex completely positive for each $k \in \{1,\dots,T\}$*, and the numbers* $\{m_\nu\}$ *satisfy the linear consistency identities*

$$\sum_{j\in Y} m_{\nu+e_j} = m_\nu \qquad \forall ||\nu||_1 \leqslant T-1.$$

Proof. (1) $\Rightarrow$ (2). Assume there exists a probability measure μ on $\Delta(Y)$ such that

$$m_\nu \;=\; \int_{\Delta(Y)} \prod_{i\in Y} \theta_i^{\nu_i} \mathrm{d}\mu(\theta) \qquad \forall ||\nu||_1 \leqslant T.$$

Linear consistency. Fix ν with $||\nu||_1 \leqslant T-1$. Then

$$\sum_{j\in Y} m_{\nu+e_j} \;=\; \int_{\Delta(Y)} \prod_{i\in Y} \theta_i^{\nu_i} \Big(\sum_{j\in Y} \theta_j\Big)\, \mathrm{d}\mu(\theta).$$

Since $\theta \in \Delta(Y)$, we have $\sum_{j\in Y}\theta_j = 1$, so

$$\sum_{j\in Y} m_{\nu+e_j} \;=\; \int_{\Delta(Y)} \prod_{i\in Y} \theta_i^{\nu_i}\, \mathrm{d}\mu(\theta) \;=\; m_\nu,$$

establishing the linear consistency identities.

Simplex complete positivity of $M^{(k)}$*.* For each $k \in \{1,\dots,T\}$, we have

$$M^{(k)}_{i_1\cdots i_k} \;=\; \int_{\Delta(Y)} \prod_{j=1}^{k} \theta_{i_j}\, \mathrm{d}\mu(\theta) \qquad \forall (i_1,\dots,i_k) \in Y^k.$$

Equivalently,

$$M^{(k)} \;=\; \int_{\Delta(Y)} \theta^{\otimes k}\, \mathrm{d}\mu(\theta).$$

The map $\theta \mapsto \theta^{\otimes k}$ is continuous and $\Delta(Y)$ is compact, so

$$M^{(k)} \in \mathrm{conv}\{\theta^{\otimes k} : \theta \in \Delta(Y)\}.$$

By Carathéodory's theorem there exist points $\pi^{(1)}, \dots, \pi^{(r)} \in \Delta(Y)$ and weights $\gamma_1, \dots, \gamma_r > 0$ with $\sum_{s=1}^{r} \gamma_s = 1$ such that

$$M^{(k)} \;=\; \sum_{s=1}^{r} \gamma_s \, \pi^{(s)\otimes k}.$$

This is exactly the definition of simplex complete positivity of $M^{(k)}$, so (2) holds.

(2) $\Rightarrow$ (1). Assume that for every $||\nu||_1 \leqslant T-1$, $\sum_{j\in Y} m_{\nu+e_j} = m_\nu$; for each $k \in \{1, \dots, T\}$, the order–k tensor $M^{(k)}$ with entries $M^{(k)}_{i_1 \cdots i_k} = m_{e_{i_1}+\cdots+e_{i_k}}$ is simplex completely positive.

We will show there exists a probability measure μ on $\Delta(Y)$ such that $m_\nu = \int_{\Delta(Y)} \prod_{i\in Y} \theta_i^{\nu_i} \, \mathrm{d}\mu(\theta)$ for all $||\nu||_1 \leqslant T$.

Step 1: A linear functional on polynomials. Let $\mathcal{A}_T$ be the real vector space of all polynomials in $(\theta_1, \dots, \theta_n)$ of total degree at most T. Define a linear functional $L : \mathcal{A}_T \to \mathbb{R}$ by

$$L\left(\prod_{i\in Y} \theta_i^{\nu_i}\right) := m_\nu \qquad \text{for all } \nu \text{ with } ||\nu||_1 \leqslant T,$$

and extend linearly to arbitrary polynomials $p(\theta) = \sum_{||\nu||_1 \leqslant T} a_\nu \prod_{i\in Y} \theta_i^{\nu_i}$ by $L(p) := \sum_{||\nu||_1 \leqslant T} a_\nu m_\nu$.

Let

$$S(\theta) := \sum_{i\in Y} \theta_i.$$

Step 2: L treats S as 1. We claim that for every polynomial $p \in \mathcal{A}_T$ with $\deg p \leqslant T-1$,

$$L(S\,p) = L(p). \tag{34}$$

It suffices to check this on monomials and extend by linearity.

Fix ν with $||\nu||_1 \leqslant T-1$. Then

$$S(\theta) \prod_{i\in Y} \theta_i^{\nu_i} = \Big(\sum_{j\in Y} \theta_j\Big) \prod_{i\in Y} \theta_i^{\nu_i} = \sum_{j\in Y} \prod_{i\in Y} \theta_i^{(\nu+e_j)_i},$$

$$L(S\prod_{i\in Y}\theta_i^{\nu_i}) = \sum_{j\in Y} L(\prod_{i\in Y}\theta_i^{(\nu+e_j)_i}) = \sum_{j\in Y} m_{\nu+e_j}.$$

By the linear consistency identities, $\sum_{j\in Y} m_{\nu+e_j} = m_\nu = L(\theta^\nu)$, so $L(S\,\theta^\nu) = L(\theta^\nu)$. By linearity, (34) holds for all p with $\deg p \leqslant T-1$.

Iterating this identity, we obtain

$$L(S^k p) = L(p) \quad \text{whenever } \deg p + k \leqslant T. \tag{35}$$

Step 3: L is nonnegative on polynomials nonnegative on the simplex. Let $p \in \mathcal{A}_T$ satisfy $p(\theta) \geqslant 0$ for all $\theta \in \Delta(Y)$. Let $d = \deg p \leqslant T$ and set $k := T - d \geqslant 0$. Define

$$\tilde{p}(\theta) := S(\theta)^k\, p(\theta),$$

which has degree at most T. By (35), $L(\tilde{p}) = L(S^k p) = L(p)$, so it suffices to show $L(\tilde{p}) \geqslant 0$. Write $\tilde{p}(\theta) = \sum_{||\nu||_1 \leqslant T} a_\nu\, \theta^\nu$. Applying (35) to each monomial θ^ν with $||\nu||_1 < T$ gives $L(\theta^\nu) = L\big(S^{T-||\nu||_1}\theta^\nu\big)$. Expanding $S^{T-||\nu||_1}\theta^\nu = \big(\sum_{i\in Y}\theta_i\big)^{T-||\nu||_1}\theta^\nu$ using the multinomial theorem yields

$$S^{T-||\nu||_1}\theta^\nu = \sum_{|\beta|=T-||\nu||_1} \binom{T-||\nu||_1}{\beta}\theta^{\nu+\beta} = \sum_{||\eta||_1=T} c_\eta^{(\nu)}\,\theta^\eta,$$

where $c_\eta^{(\nu)} = \binom{T-||\nu||_1}{\eta-\nu}$ if $\eta \geqslant \nu$ componentwise and $c_\eta^{(\nu)} = 0$ otherwise. Applying L gives $L(\theta^\nu) = \sum_{||\eta||_1=T} c_\eta^{(\nu)}\, m_\eta$, which expresses $L(\theta^\nu)$ entirely in terms of the degree-T moments $\{m_\eta\}_{||\eta||_1=T}$. Therefore, by linearity of L,

$$L(\tilde{p}) = \sum_{||\nu||_1\leqslant T} a_\nu\, L(\theta^\nu) = \sum_{||\eta||_1=T} b_\eta\, m_\eta = \sum_{||\eta||_1=T} b_\eta \sum_{s=1}^{r} \gamma_s\,(\pi^{(s)})^\eta = \sum_{s=1}^{r}\gamma_s \sum_{||\eta||_1=T} b_\eta\,(\pi^{(s)})^\eta = \sum_{s=1}^{r}\gamma_s\,\tilde{p}\big(\pi^{(s)}\big),$$

where $b_\eta := \sum_{||\nu||_1\leqslant T} a_\nu\, c_\eta^{(\nu)}$, and the third equality uses simplex complete positivity of $M^{(T)}$. The last equality holds because the reduction via $S = 1$ ensures that $\sum_{||\eta||_1=T} b_\eta\,(\pi^{(s)})^\eta = \tilde{p}(\pi^{(s)})$ for each $\pi^{(s)} \in \Delta(Y)$. Since each $\pi^{(s)} \in \Delta(Y)$, we have $S(\pi^{(s)}) = 1$ and hence $\tilde{p}(\pi^{(s)}) = p(\pi^{(s)}) \geqslant 0$. Therefore $L(\tilde{p}) \geqslant 0$, and thus

$$L(p) \geqslant 0 \quad \text{whenever } p \in \mathcal{A}_T \text{ and } p(\theta) \geqslant 0\ \forall\,\theta \in \Delta(Y). \tag{36}$$

Step 4: m lies in the convex hull of truncated moment vectors. Let $\mathcal{I} := \{\nu \in \mathbb{N}^n :$

$||\nu||_1 \leqslant T\}$, and let $m := (m_\nu)_{\nu\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$. For each $\theta \in \Delta(Y)$, define the truncated moment vector

$$\phi(\theta) := (\theta^\nu)_{\nu\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}.$$

Let

$$C := \operatorname{conv}\{\phi(\theta) : \theta \in \Delta(Y)\}$$

be the convex hull of all such vectors. We claim $m \in C$. Suppose, to the contrary, that $m \notin C$. Since C is a compact convex subset [14] of the finite–dimensional space $\mathbb{R}^{\mathcal{I}}$, the separating hyperplane theorem implies that there exist a nonzero vector $a = (a_\nu)_{\nu\in\mathcal{I}}$ and a scalar α such that

$$\sum_{\nu\in\mathcal{I}} a_\nu m_\nu \;<\; \alpha \quad \text{and} \quad \sum_{\nu\in\mathcal{I}} a_\nu \phi(\theta)_\nu \;\geqslant\; \alpha \quad \forall\, \theta \in \Delta(Y).$$

Define the polynomial

$$p(\theta) := \sum_{\nu\in\mathcal{I}} a_\nu\, \theta^\nu - \alpha.$$

Then $p(\theta) \geqslant 0$ for all $\theta \in \Delta(Y)$ by construction, while

$$L(p) = \sum_{\nu} a_\nu m_\nu - \alpha \;<\; 0.$$

This contradicts (36). Hence $m \in C$.

Step 5: Constructing a representing measure. Since $m \in C$, there exist $\theta^{(1)}, \ldots, \theta^{(r)} \in \Delta(Y)$ and weights $\lambda_1, \ldots, \lambda_r \geqslant 0$ with $\sum_{s=1}^r \lambda_s = 1$ such that

$$m_\nu \;=\; \sum_{s=1}^{r} \lambda_s\, (\theta^{(s)})^\nu \qquad \forall\, \nu \in \mathcal{I}.$$

Define a probability measure μ on $\Delta(Y)$ by

$$\mu := \sum_{s=1}^{r} \lambda_s\, \delta_{\theta^{(s)}},$$

[14] Since $\Delta(Y)$ is compact and the map $\theta \mapsto \phi(\theta) = (\theta^\nu)_{\nu\in\mathcal{I}}$ is continuous, its image is compact, and therefore C is compact as a convex hull of a compact set in the finite-dimensional space $\mathbb{R}^{\mathcal{I}}$.

where $\delta_{\theta^{(s)}}$ is the Dirac measure at $\theta^{(s)}$. Then, for all $||\nu||_1 \leqslant T$,

$$\int_{\Delta(Y)} \theta^\nu \, \mathrm{d}\mu(\theta) = \sum_{s=1}^{r} \lambda_s \, (\theta^{(s)})^\nu = m_\nu.$$

Thus μ is a probability measure on $\Delta(Y)$ with the required moments, establishing (1).

This completes the proof of the equivalence of (1) and (2). □

Lemma B.3. *For any horizon $T \geqslant 0$, $\sum_{(i_1,\ldots,i_T)\in Y^T} M^{(T)}_{i_1\cdots i_T}(\mathbf{q}) = 1$.*

Proof. We proceed by induction on T.

Base case ($T = 0$). By definition $M^{(0)}(\mathbf{q}) = m_{\mathbf{0}} = 1$.

Inductive step. Suppose the claim holds for T. Let $\nu = e_{i_1} + \cdots + e_{i_T}$. Then

$$\begin{aligned}
\sum_{(i_1,\ldots,i_{T+1})\in Y^{T+1}} M^{(T+1)}_{i_1\cdots i_{T+1}}(\mathbf{q}) &= \sum_{(i_1,\ldots,i_{T+1})\in Y^{T+1}} m_{e_{i_1}+\cdots+e_{i_{T+1}}} = \sum_{(i_1,\ldots,i_T)\in Y^T} \sum_{i_{T+1}\in Y} m_{\nu+e_{i_{T+1}}} \\
&= \sum_{(i_1,\ldots,i_T)\in Y^T} \sum_{i_{T+1}\in Y} q^{(\nu)}_{i_{T+1}} \, m_\nu = \sum_{(i_1,\ldots,i_T)\in Y^T} m_\nu \underbrace{\sum_{i_{T+1}\in Y} q^{(\nu)}_{i_{T+1}}}_{=1} \\
&= \sum_{(i_1,\ldots,i_T)\in Y^T} M^{(T)}_{i_1\cdots i_T}(\mathbf{q}) \;=\; 1.
\end{aligned}$$

where the first and fifth equalities use the definition of the count moment tensor, the third uses the recursion $m_{\nu+e_i} = q^{(\nu)}_i m_\nu$, the fourth uses $q^{(\nu)} \in \Delta(Y)$, and the last is the inductive hypothesis. □

Lemma B.4. *If the count moment tensor $M^{(T)}(\mathbf{q})$ is completely positive, then it is simplex completely positive.*

Proof. Since $M^{(T)}(\mathbf{q})$ is completely positive, there exist a positive integer r and nonnegative vectors $\bar{\pi}^{(1)}, \ldots, \bar{\pi}^{(r)} \in \mathbb{R}^Y_{\geqslant 0}$ such that

$$M^{(T)}_{i_1\cdots i_T}(\mathbf{q}) = \sum_{s=1}^{r} \bar{\pi}^{(s)}_{i_1} \cdots \bar{\pi}^{(s)}_{i_T}.$$

We may assume each $\bar{\pi}^{(s)}$ is nonzero. For each s, define

$$Z_s := \sum_{i \in Y} \bar{\pi}_i^{(s)} \;>\; 0, \qquad \pi_i^{(s)} := \frac{\bar{\pi}_i^{(s)}}{Z_s}, \qquad \gamma_s := Z_s^T.$$

Then $\pi^{(s)} \in \Delta(Y)$ since $\pi_i^{(s)} \geqslant 0$ and $\sum_{i \in Y} \pi_i^{(s)} = 1$. Substituting $\bar{\pi}_i^{(s)} = Z_s \pi_i^{(s)}$ gives

$$M^{(T)}_{i_1 \cdots i_T}(\mathbf{q}) = \sum_{s=1}^{r} \gamma_s \, \pi_{i_1}^{(s)} \cdots \pi_{i_T}^{(s)}.$$

It remains to verify $\sum_{s=1}^{r} \gamma_s = 1$. Summing the decomposition over all $(i_1, \ldots, i_T) \in Y^T$ and using $\pi^{(s)} \in \Delta(Y)$:

$$\sum_{(i_1,\ldots,i_T) \in Y^T} M^{(T)}_{i_1 \cdots i_T}(\mathbf{q}) = \sum_{s=1}^{r} \gamma_s \underbrace{\sum_{(i_1,\ldots,i_T) \in Y^T} \pi_{i_1}^{(s)} \cdots \pi_{i_T}^{(s)}}_{= \left(\sum_{i \in Y} \pi_i^{(s)}\right)^T = 1} = \sum_{s=1}^{r} \gamma_s.$$

By Lemma B.3, the left-hand side equals 1, so $\sum_{s=1}^{r} \gamma_s = 1$. Therefore $M^{(T)}(\mathbf{q})$ is simplex completely positive. □